

Test items and translation: capturing early conceptual development in mathematics reliably?

Abstract

Translating items of educational tests from one language to another is problematic. Arriving at accurate translations of concepts formulated in a language that is grammatically and syntactically incommensurable with a target language is a concern that probably won't find resolution. And the very act of translation can obscure the accuracy of test items. Item Response Theory holds that only the ability of people tested and the difficulty of items should have an effect on the dataset. We report on an instance where a test was translated from German to English and then into isiZulu and Sesotho. We tested 106 pupils from similar socio economic backgrounds and schools. Our aim is to determine whether the translation had any effect on the accuracy of the instrument, which has been normed and standardised in Germany with a sample of over 1000 grade 1 pupils.

Keywords: Item response theory, conceptual development, mathematical competence, learning and assessment.

Graham Dampier, University of Johannesburg. E-mail: gadampier@uj.ac.za. Daphney Mawila, University of Johannesburg. E-mail: dmawila@uj.ac.za

South African Journal of Childhood Education | 2012 2(2): 35-57 | ISSN: 2223-7674 | © UJ

Introduction: signs and signified in the translation of a test

Translated items of educational diagnostic tests are likely to come with a host of possible problems. In any translation these may extend beyond the issues associated with trying to arrive at an approximate equation of meaning from the original to the translated text. Translation, particularly in the case of languages that exhibit little commonality, is, by force, constrained. The constraint is due not only to disparate representational systems, each with their own grammar, syntax and lexicon, but more problematically to pre-linguistic epistemological stores that are determined and defined by different geo-cultural points of reference that are captured in different languages as systems of signs (Benjamin, 1997; Jakobson, 1959; Gopnik and Meltzoff, 1997). Any given language is involved in the continued production and naming of new concepts that will in the end be expressed through its system of signs (Evans & Levinson, 2009; Levinson, 2003). Simply put, different languages tend to represent concepts and ideas that are often unique to the environment in which they exist. An often-used example comes from the Inuit language and its variants, which have many different words for ‘snow’ that describe variation within the phenomenon. Also, ‘Nguni languages have the same word for ‘green’ and ‘blue.’

However, research on what Carey (1985, 2009) has referred to as “core cognition” suggests that children from disparate language groups share certain innate fundamental predispositions that form the cognitive basis for developing concepts, or, to be more precise, new and more complex *conceptual representations*. A Whorfian (Whorf, 1956) stance does not explain this type of content architecture of cognition, but does have much to say for the signs by which cognition is communicated and which develop uniquely in speech communities. In formal education, if one uses De Saussurean parlance (De Saussure, 1986) the *signs* of language are as important as their *signifieds* (Nazzi & Gopnik, 2001).

The larger topic of how language feeds into education, and vice versa, is not the focus of this article. We address the way specific test items are used to assess children, by arguing that the construct validity of these items rests on consensus of the meaning of test items, across the four languages we will refer to. We describe an example of the attempt to translate a latent trait encoded in one test from German into English and then into Sesotho and isiZulu. The MARKO-D, an instrument that assesses mathematical competence and which was developed in Germany (Ricken, Fritz & Balzer, 2012), is a test that scratches the surface of basic mathematical cognition. The test itself is, however, aimed at capturing learned representations.¹

The test was translated into English (originally not the South African variant) and from there into isiZulu and Sesotho. Some of the formal isiZulu and Sesotho versions

1 There are exceptions. Cultures, such as the Piraha and Mundurucu in the Amazon, have not developed the ability to go beyond core cognition of number, due merely to a lack of need. These cultural systems do not rely on numerical representations that have developed beyond what evolution has endowed the human species with. In these languages only the first three numbers are named, while the 1-1 correspondence of objects and cardinal quantities has not been established for more than three objects. (Carey, 2009; p. 303)

of the test items were virtually incomprehensible to the grade 1 pupils it sought to assess empirically, since they were represented in the formal register of the two languages that the urban children do not use (Welch, Dampier and Mawila, 2011). As a result, the validity and reliability of the test and the data it sought to gather could have been compromised, since the representational mode in which the test was delivered appeared, on face value, not to suit the linguistic context in which it was administered.

The purpose of this article is two-fold. Firstly, using the *MARKO-D* test, we evaluate the cross-linguistic assessment of mathematical representations built on core mathematical cognition. The developers of the test explain that it was designed for pre-school children (Gerlach, Fritz & Leutner, 2012) and was piloted and later validated with 3000 children in Germany. They explain that it was developed not to test core knowledge of mathematical concepts (which is believed to be innate), even though it taps into this, but that it was designed to measure mathematical concepts that have been learned socially, from the ability to count with the succession principle through to knowledge of natural number and cardinality. One of the basic assumptions of the test is that while all children are born with a set of core mathematical concepts, they develop basic mathematical skills as they learn, mostly through some instruction, be it formal or informal, which is mediated, mostly, by language as semiotic mediator.

Secondly, we assess whether the double translation from German into English and then into isiZulu and Sesotho had any effect on the way in which individuals from the different language groups responded to the test. The Rasch model of measurement (Rasch, 1960; Andrich, 1988; Bond & Fox, 2007) was used to assess the effect of the double translation on the items of the test. The model indicates whether the *MARKO-D* is a reliable tool with which to measure mathematical competence in English, isiZulu and Sesotho in urban South African classrooms. It was used to determine whether our data violated the assumption of *unidimensionality*, which is a basic tenet of the Rasch model and item response theory, more generally (Wright, 1988; Wright & Stone, 1999, Andrich 1988; de la Torre & Patz, 2005; Andrich & Hagquist, 2012; Bond & Fox, 2007; Rasch, 1960; DeMars, 2010).

This article presents a preliminary analysis of the suitability of the *MARKO-D* as a measure of mathematical competence in what is a very distinct setting, the urban South African ‘township’ classroom. The sample size of the pilot test on which we report ($n=106$) is suitable for such a pilot and was followed in 2012 by a sample of 280. De Mars explains that (2010, p. 34), “Rasch or one-parameter logistic (1PL) models are often used with samples as small as 100 or 200 examinees,” but prevailing wisdom holds that increasing the number of examinees will improve the estimation of person ability and item difficulty. With the limitations of this sample size, we make no claim to have arrived at a conclusive evaluation of the accuracy, validity and reliability of the *MARKO-D* for all urban South African schools that are similar to our sample school. Instead, this analysis of the test provides a preliminary assessment of the suitability of the test and the effect of translating the test from German to English and then into isiZulu and Sesotho.

The Rasch model indicates (1) whether the MARKO-D assesses mathematical competence consistently in the three different language groups, in other words, it determines to what extent the test fits with the population assessed in this study, and (2) whether the act of translating the test into three different representational systems has had any effect on the dimensionality of the data. The principle of *unidimensionality* maintains that only the difficulty of test items and the ability of individuals should have any bearing on the data (Wright, 1998; Wright & Stone, 1999, Anrdrich 1988; Andrich & Hagquist, 2012; Bond & Fox, 2007, DeMars, 2010, de la Torre & Patz, 2005). Violation of this principle may suggest that extraneous variables are clouding the accuracy of the MARKO-D as a measure of mathematical competence.

We begin with a brief discussion of what core mathematical cognition refers to, and, specifically, what Carey's (2009) theoretical stance entails, before exploring her theory of learning, or cognitive development. As a leading cognitive developmental psychologist, her view is widely accepted (Carey, 2001; 2004 & 2009). This will be followed by a brief discussion of the origin and the validation of the MARKO-D test. Thereafter, we report on how we used the Rasch model to assess whether this instrument accesses mathematical competence consistently and accurately across the three language groups tested. We also discuss the effect that translating the assessment has had on the different language groups.

Core mathematical cognition

Carey (2009, p. 3) explains that her work, *The Origin of Concepts*, offers an account of the "human capacity for conceptual representation." On the surface her argument appears to be rather simple, since her basic premise is that (2009, p. 3), "(s)ome concepts, such as *object* and *number*, arise in some form over evolutionary time," while other concepts "such as *kayak*, *fraction*, and *gene*, spring from human cultures, and the construction process must be understood in terms of both human individuals' learning mechanisms and sociocultural processes."

The distinction between what is inherent in our species (produced over the course of our evolution) and what cultures produce in the attempt to articulate what is particular to them, is maintained throughout the exploration of what she refers to as conceptual *representation* and conceptual development. Carey's argument for core cognition has been consistent for over a decade.

Along with Spelke (Carey and Spelke, 1994, p. 169), she argues that, "human reasoning is guided by a collection of innate domain specific systems of knowledge" and that each system is defined by "a set of core principles that define the entities covered by the domain and support reasoning about those entities." Learning, they argue, will entail the development and the entrenchment of these core principles (1994, p. 169). The domains of knowledge they are referring to include knowledge of language, knowledge of physical objects and knowledge of number (Chomsky,

1980; Carey & Spelke, 1994)². Later, Carey & Spelke (1996, p. 517) argue that in the same way we “are endowed with multiple, specialized perceptual systems, so we are endowed with multiple systems for representing and reasoning about entities of different kinds,” which includes “at least four core conceptual systems encompassing knowledge of objects, agents, number, and space.”

While Carey has since distanced herself from the “core knowledge” argument forwarded by herself in collaboration with Spelke, and by other cognitive developmentalists (Carey & Spelke, 1996; Spelke, 2000; Spelke, 2003; Gelman, 1991; and Leslie, 1994³), her definition of “core cognition” retains conceptual and theoretical relations to it. She (2009, p. 10-11) maintains the stance that cognitive development is organized around a core set of conceptual representations that are “evolutionarily underwritten,” and that differ from sensory and perceptual representations “in having a rich, conceptual, inferential role to play in thought.” The major difference between core cognition and core knowledge is that core cognition, notably, does not claim to be veridical, which means that core cognition exists independently of contact with particular phenomena in the world. Core cognition is embedded within the brain, it exists as the *fons et origo* of knowledge, and does not require experience to actively serve its basic function, which is to make sense of sensory input. Even though knowledge may change over time, it appears that core cognition remains stable and constant.⁴

By implication members of different language groups, for example the German, isiZulu, English and Sesotho groups, should all, at some point of maturation, have similar, if not the same, set of core mathematical concepts available to them before they embark on the process of creating new representational systems that will arise through sociocultural “semiotic mediation” (Vygotsky, 1978). Cognitive neuropsychologists have gone a long way to proving that all human beings, regardless of language and culture, and indeed many animals such as rats, the various primate species, many species of bird, and so on, are naturally endowed with the ability to analyse reality in terms of number and quantity (Carey, 2009; Dehaene, 2010; Deheane & Brannon, 2011; Feigenson, Dehaene & Spelke, 2004; Feigenson, Carey & Spelke, 2002; Le Corre & Carey, 2007). If core cognition is an evolutionary product, then every child, regardless of linguistic origin, should have the same basis from which to develop new and more sophisticated concepts, theories and systems of representation.

2 Leslie (1994, p. 121) writes that: “The language faculty is probably not the only member of a class of core domains concerned with knowledge of formal systems. Formal core domains plausibly include *number* and *music*, as well as *grammar*.”

3 Leslie (1994) prefers “core architecture.”

4 Carey & Spelke (1996, p. 519) argue that, “[c]ore systems, in contrast, are elaborated but not revised: neither infants, nor children, nor adults engaged in commonsense reasoning ever give up their initial systems of knowledge.” It is foundational in humans as it is foundational in other species.

‘Bootstrapping’ and the representation of positive integers

“Bootstrapping is the process that underlies the creation of such new concepts and thus it is part of the answer to the question: What is the origin of concepts?”

(Carey, 2004: p. 60)

In this section of the article we explore the debate around the continuity of core number cognition and its role in learning. There are two trends in the literature, seeking to explain the genesis of number and other mathematical concepts.

Carey (2009, p. 12) explains that *The Origin of Concepts* is an “extended exploration of learning mechanisms,” and she characterizes “learning processes as those that build representations of the world on the basis of computations on input that is itself representational.” The world is understood through representational input, which can be observed or experienced, that will ultimately be encoded in a conceptual system. An elemental cistern of concepts serves as the foundation for learning and development. Carey explains (2009, p. 13) that, the “proposals for the initial stock of representational primitives and for the types of learning mechanisms that underlie cognitive development are logically independent, although there are inductive relations between them.” This posits that learning is the reconciliation of new information and pre-existing (innate) cognition, which is conducted by inherent learning mechanisms and the representational systems (such as language) in which they are communicated. Language as semiotic mediator is plainly crucial.

The development of a mathematical representational system is an example of this. While individuals share a core set of mathematical concepts that are the product of human evolution, each individual will create a unique mathematical representational system as he/she is exposed to further cultural constructions of mathematics. Carey writes (2009, p. 18) that the “integer list is a cultural construction with more representational power than any of the core representational systems on which it is built.” This means that integers, as products of conceptual development and cultural construction, are built from more basic forms of mathematical cognition. Similarly, the development of rational numbers “transcends the representations available at the outset of [this particular] construction process,” which, according to Carey (2009, p 19), includes the “representation of integers created by children during their preschool years.”

Children with different linguistic and cultural origins will develop in line with the proclivities and limitations of the system within which they develop. Since some learning, be it formal or informal would have occurred before the children tested in this study entered school, language (as a system of signs that signify concepts) would have played a major role in their development. The ability of these children to represent positive integers and to understand cardinality is defined by the language they speak and the culture they come from. Here core cognition is not the handmaiden of learning – now language and environment kick in.

Carey argues that while analog magnitude representations of number, the parallel individuation system and natural language quantifiers, are all products of human

evolution (the last being a product of language evolution in particular), and qualify as examples of “core cognition,” the ability to represent natural numbers, or positive integers, is the product of socio-cultural mediation.⁵ The human ability to represent positive integers is learned. It is a product of culture, not nature. She explains this, with the notion of conceptual systems (CS) that change from one system to another (2009, p. 292):

We can examine the core cognition systems with numerical content and specify the ways in which integer representations transcend them. The most important task to one attempting to establish conceptual discontinuity is to characterize CS1 and CS2, demonstrating in what sense CS2 contains representations not expressible in CS1.

Since the representations of natural numbers are learned, they form part of a conceptual system, in this case CS2, that is conceptually discontinuous with core cognition system that 1) analog magnitude representations of number, 2) the parallel individuation system and 3) natural language quantifiers are part of, which is represented here as CS1 (see figure 1).

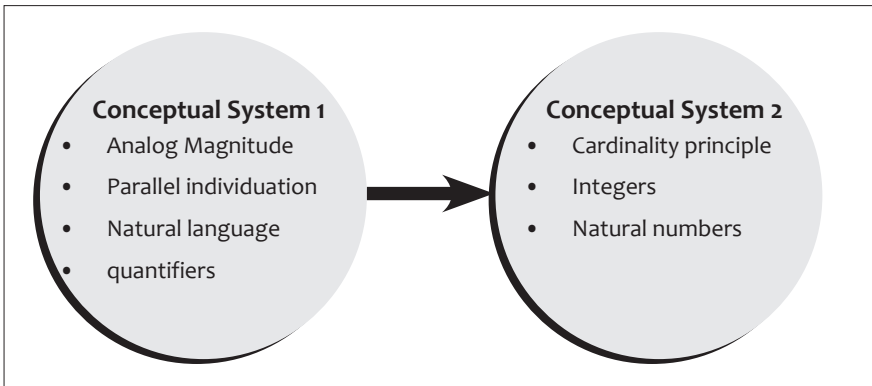


Figure 1: The conceptual representation of integers.

Since the publication of Gelman & Gallistel’s *The Child’s Understanding of Number* (1978) cognitive psychologists have forwarded conflicting theories of how children learn to “determine the cardinality of a collection” (Rips, Bloomfield, & Asmuth, 2008, p. 626; see also Le Corre & Carey, 2007 and Le Corre, Van de Walle, Brannon, & Carey, 2006). Many cognitive psychologists have argued that two innate conceptual systems of number form the basis for determining cardinality when children are roughly between 4 and 5 years old (Dehaene, 1997; Feigenson, Dehaene, & Spelke, 2004; Starkey, Spelke & Gelman, 1990; Holloway & Ansari, 2009; and by Carey herself in various works). There is evidence that children as young as six months of age are capable of producing analog magnitude representations of number and parallel individuations of small sets. These two innate systems of numerical representation, as well as the quantifiers of

5 Elsewhere she argues (2004, p. 60): “Before they acquire language, infants form several different types of representation with numerical content, at least two of which they share with other vertebrate animals.”

natural language, have been cited by many, Gelman & Gallistel (1978) most notably, as forming the basis for children’s ability to represent positive integers.

However, Carey and other authors (2004; 2009; Le Corre & Carey, 2007; Le Corre et al, 2006) present sceptical views of the plausibility of the argument that these three innate systems of number are capable of representing positive integers and allowing children to arrive at an understanding of cardinality. Le Corre & Carey (2007) explain that, firstly, the *analog magnitude system* is not capable of presenting the exact quantities of number that the representation of positive integers requires. The analogue magnitude system cannot represent numbers as discrete quantities, because it serves only to estimate quantity and to distinguish between magnitudes of differing size.⁶ The analogue magnitude system “encodes number as would a number line,” and it is present in human beings as early as “the sixth month of life” (ibid, 2007, p. 397). Le Corre & Carey conclude that the analogue magnitude system is too limited to serve as a plausible basis for developing the ability to represent positive integers (2007, pp. 433-434). It is not capable of dealing with the precise values of number required by the representation of positive integers.

Secondly, these authors also argue that (ibid, pp. 297-298) the *parallel individuation system* is capable of representing “sets of individuals by creating working memory models in which each individual in a set is represented by a unique mental symbol.” They are quick to caution, however, that this “system has a hard capacity limit” (ibid, p. 298). They explain:

In adults, it cannot hold any more than 4 individuals in parallel. Many experiments suggest that the infant system cannot hold any more than 3 individuals in parallel, though one group of researchers has found that it too can hold up to 4. Importantly, unlike the analog magnitude system, this system contains no symbols for number. However, it is clear that it has numerical content. Criteria for numerical identity (sameness in the sense of same one) determine whether a new symbol is created in a given model. Additionally, infants can create working memory models of at least two sets of 3 or fewer individuals, and can compare these models on the basis of 1–1 correspondence to determine numerical equivalence or numerical order.

Thirdly, they argue, the *quantifiers of natural language* are similarly limited in their capacity to represent discrete numerical quantities. Le Corre & Carey (2007, p. 298) explain:

A third system available to non-linguistic primates and to preverbal infants is what we will call the “set based quantificational system.” This system is the root of the meanings of all natural language quantifiers. To provide the basis for quantification, this system explicitly distinguishes the atoms, or individuals, in a domain of discourse from all the sets that can be comprised of them.

⁶ Le Corre & Carey (2007, p. 397) write that the analogue magnitude “is characterized by two related psychophysical signatures –Weber’s law and scalar variability. Weber’s law states that discriminability of two quantities is a function of their ratio (e.g. 5 and 10 are easier to discriminate than 45 and 50). Scalar variability holds if the standard deviation of the estimate of some quantity is a linear function of its absolute value. Analogue magnitude serves to estimate various quantifiable areas of experience (2004, p. 60): “Mental analog magnitudes represent many dimensions of experience - for example, brightness, loudness, and temporal duration. In each case as the physical magnitudes get bigger, it becomes increasingly harder to discriminate between pairs of values that are separated by the same absolute difference.”

The representation of natural numbers, via numeral (counting) lists, depends on the ability to 1) count in sequence and 2) the accompanying ability to associate these representations with discrete quantities that follow successively, which equates to $n + 1$ for any number in the count sequence. These abilities are both not innate but learnt. This requires the formation of a direct relation between, for instance, the representation “five” and the quantity of five, represented here as five lines “- - - -”. The ability to do this, according to Carey (2009, p. 295), far exceeds the limitations of analog magnitude representations of number, which she explains are not “powerful enough to represent the natural numbers and their key property of discrete infinity.” What is more, analog magnitude representations embedded within CS1 “provide merely approximate representations of the numbers in their domain, even one, whereas numeral list systems [which are part of the CS2] represent each natural number exactly.”

She finds (2009, p. 295) that while systems of parallel individuation may contain “machinery for indexing and tracking sets of individuals,” it does not contain “symbols for cardinal values.” Indicating that the ability to represent natural number exceeds the conceptual limit of parallel individuation systems. In addition, “natural language includes no representations of exact cardinal values above three” (Carey, 2009, p. 296).

At this stage the question is: “How do children learn to represent positive integers, that is, are there stages in the process of gaining knowledge of cardinality?”

Carey’s model of learning rests on a characterisation of learning via a process she calls “bootstrapping” – a metaphor with a specific history, which we will not explore in this article. While there is not enough space to explore this model of learning in the detail it merits, it is expedient to provide a brief explanation for this specific use of term.⁷ Since Carey’s model of conceptual change requires that conceptual discontinuities exist when new representational systems are developed, the metaphor of bootstrapping maintains that learning entails “building a structure while it is not [already] grounded” (2009, p. 306). Initially, when new representational systems are developed, the learner only learns “the relations of a system of symbols to one another,” and does not associate these symbols with pre-existing concepts (2009, p. 306). This process of association is described as mapping each sign onto a concept or a component of a concept. Before this mapping takes place, signs serve only as *placeholders*, that is, they are spectres of the concepts they will at some point serve to represent in a direct and semantically constrained relation. Once the signs and concepts of a new representational system correspond to one another, the learner has successfully learned how to associate a new system of signs with the concepts they serve to represent. A new conceptual system has hereby been developed. Once learners have grasped the notion of ‘decades’ and ‘decoupling’, they cannot go back to a conceptual system (CS1) in which this understanding (CS2) does not feature.

7 For evaluations of “bootstrapping” as a model of learning see Rips, Asmuth & Bloomfield (2008; 2006); Gelman & Butterworth (2005); and Butterworth (2005) - for criticism of the approach, and Carey (2004 and 2009), Henning (2012) and Henning & Dampier (2012) - for a description of the process.

We can explain the notion of place holding further. In learning how to represent integers or natural numbers, children at first learn the signs of representation before they come to understand the cardinality of number. And, it appears, that process of learning occurs in stages (2009, p. 301):

Besides confirming that CS1 and CS2 are stable representational systems, these data constrain [sic - contain] an account of the learning process. After having memorized the count list and the count routine [which constitute the signs of the new representational system], first the child is a no numeral knower, although by 24 months of age, many English learning children are already “one”-knowers. Being a “one”-knower is a consistent stage children remain in for six to nine months. Then they become “two”-knowers, and some also become “four”-knowers before figuring out how the numeral list represents natural number. Children stay subset-knowers for 1 to 1.5 years. When they become cardinal principle knowers, they have created a representation of some positive integers, a numerical representation [sic] that transcends core number representations.

The bootstrapping process, through which the representation of positive integers is learned, requires, at first, that the learner learn the system of signs with which natural numbers will be represented. At first these signs are spectres, since they are devoid of any conceptual content and exist without stable signifieds. This changes once a direct relationship between the sign and the concept is established; at this point the signs are no longer placeholders or spectres for concepts they are meant to represent. In a sense, at this point the concepts or signifieds come-to-life within the sign, much as Vygotsky also theorized learning form and through words (Vygotsky, 1992; Henning, 2012).

In learning how to represent positive integers, children will go through various stages. They will begin to learn the various signifiers of the numeral list, before they understand that each sign represents a particular quantity (this encompasses the shift from “no”-knower, to “one”-knower, “two”-knower, “three”-knower, and eventually to “four”-knower, see figure 2). During the subset-knower phase, learners “know the numerical meaning of only a subset of numerals on their count list” (2009, p. 298), but do not know that, “the last word reached in a count [of a subset of the count list] represents the cardinal value of the set” (2009, p. 298). Cardinal knowers are said to understand “the numerical meaning of the activity of counting and can now reliably produce sets with the cardinal value of any numeral on their count list” (2009, p. 298).

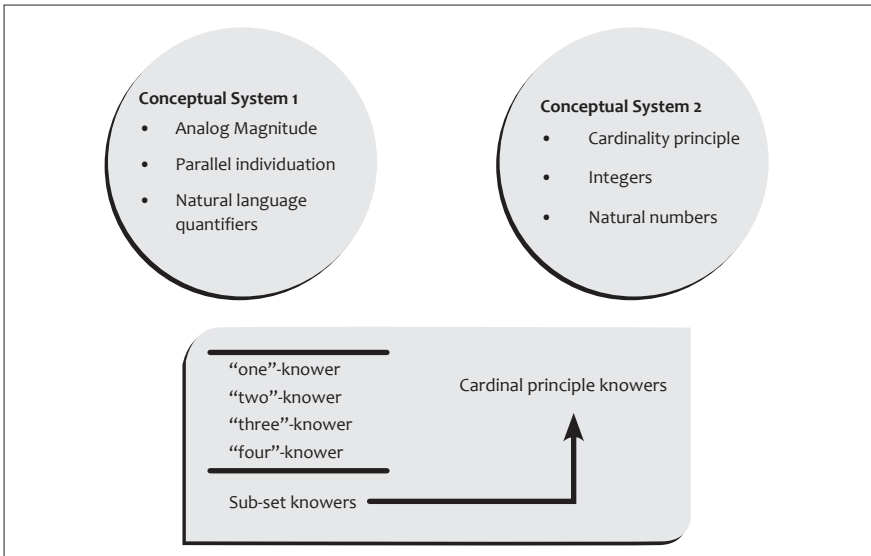


Figure 2: From “one”- knower to Cardinal principle knower .

The MARKO-D test

We conducted this pilot study at the beginning of a longitudinal panel research programme in which we follow the participants over four years in various areas of childhood development, one of which is mathematical competence. The *MARKO-D* was specifically designed to capture development of individual children and to show the way in which children develop mathematical competence. This process of development is captured in five levels that increase with complexity and difficulty. Our thesis is that every more advanced level of development is a representational system that is conceptually incommensurate with a level that has gone before. This means that more basic levels in this measure of mathematical competence cannot represent the concepts of more advanced levels, since each concept and level builds on what has come before. Ricken, Fritz-Stratmann and Balzer (2011, p. 256), the designers of the test, describe the instrument as follows (translated from German):

Process-oriented diagnostics require a theoretical framework which allows for the description and interpretation of individual competence changes. For the *MARKO-D* test, a corresponding dimension/scale of mathematical achievement was developed on the basis of theoretical suppositions and empirical data. Five concepts [which are representational, conceptual systems in Carey’s parlance] are tested, namely numbers as counting sequence, the ordinal number line, cardinal understanding, part–part–whole, and the concept of congruent intervals. There is empirical evidence for the validity of the model, using a unidimensional Rasch model. The test, on the one hand, allows for comparison of individual data with a social norm and, on the other hand, is

usable to make valid statements about individual changes and development as well as the current competence status of a child.

The MARKO-D measures children's early mathematical and arithmetical representations. While the test does not claim to access core mathematical cognition directly, it does open up empirical windows to basic conceptual representations of number, quantity, order and relationality. The test was empirically scaled, using Rasch modelling. Items that could be solved with the same conceptual knowledge were grouped together in levels that the authors refer to above. These levels of competence (broadly sequential in development) include:

1. Numerals – counting words
2. Ordinal number line
3. Cardinal numbers and partitioning
4. Conservation and class inclusion
5. Relationality

The test was standardized with data of 3000 children (4-8 years) from different contexts in Germany and it was normed with data from 1029 children (48 – 87 months), with 567 boys and 528 girls ($M = 64.6$ months; $SD = 7.2$) (Ricken et al: 2011, p.261).

To borrow from Carey, every level of competence represents a representational system that is more expressive and more complex than the level that has gone before. Every level hereby “transcends in some qualitative way” the representational system that has gone before (Carey, 2004: p 59). Each level also represents a limit that must be transcended if new and more powerful representational systems are to develop.

Administering the test in an urban South African setting

The basic premise of the Rasch model of measurement is that the ability of persons tested is independent of the difficulty of items used to measure a latent trait (Wainer, Morgan & Gustafsson, 1980; Linacre, 2002). The model is used to estimate the probability that a person with an ability of β_x will get an item with the difficulty of δ_x correct. Like other IRT models it assumes that a *single construct* or latent trait is being measured (Wainer et al, 1980; Wright, 1998; Wright & Stone, 1999; Andrich, 1988; Andrich, Marias & Humphrey, 2011; Bond & Fox, 2007, Linacre, 2002). For this reason, it makes the assumption that all items measure the same trait (Wainer et al, 1980) and that individual persons are responding to items in a way that is consistent with other people (Wright & Stone, 1999)⁸. The principle of *unidimensionality* maintains that no other dimension or extraneous variable should have any bearing on what is tested by the items, or on the way in which persons respond to items, that is, this principle

8 Wright & Stone (1999, p. 48) explain: “The Rasch model is derived from the requirement that person measures and item calibrations be separately estimable. This requires that (1) a more able person always have a greater probability of success on any item than a less able person, and (2) any person always be more likely to do better on an easier item than on a harder one.”

requires that only one construct is measured in persons and by items. Bond & Fox (2007, p. 32) explain that the “focus on one attribute or dimension at a time is referred to as *unidimensionality*.”

Unidimensionality is a general concern of item response theory (DeMars, 2010). Cook, Dorans & Eignor (1988, p. 19) argue that a basic assumption of the “most commonly used item response theory (IRT) models is that the data are unidimensional, that is, statistical dependence among item scores can be explained by a single dimension.” Even when IRT models are used to measure multiple dimensions, “a common practice in educational measurement is to estimate these abilities independently of each other” (de la Torre & Patz, 2005, p. 295). The delimitations of the Rasch model require that data reflect the measurement of one latent trait according to the ability of individual persons, and the difficulty of items, on the trait (DeMars, 2010). This means that data must fit with the assumptions and expectations of the model (Linacre, 2002).

It is assumed that all items measure the same construct or latent trait. A corollary of this is that all persons are expected to respond to the items in a way that is consistent with their ability on the trait and in relation to the ability of other persons. When these two requirements are satisfied the data are likely to fit the parameters of the model (Wright, 1998, Bond & Fox, 2007). Or as Andrich (1988, p. 61) writes “if there is discord between the data and the model, it is left open to question whether the model or the data are at fault. It is conceivable that when the data do not accord with the model with which they are intended to accord, that something has gone astray with the data or its collection.” In determining the fit of the data with the model, we are essentially trying to determine whether the data accord with the delimitations of the model. Write & Stone explain (1999, p. 47):

A complete analysis must include an evaluation of how well the data fit this essential specification. If a person answers the hard items on a test correctly but misses several easy items, we are surprised by the resulting implausible pattern of incorrect responses. While we could examine individual records item-by-item to determine this kind of invalidity, in practice we want to put such evaluations on a systematic and manageable basis. We want to be specific but also objective in our reaction to implausible and hence invalid observations.

Does it Fit?

According to Wright & Stone (1999, p. 48), “(f)it analysis shows us the extent to which any data can be used to construct measures. Each data analysis must include an evaluation of how well those particular data fit the expectations which measurement requires.” One way of determine the fit of the data to the parameters of the model is to examine the fit statistics of the data. Bond & Fox (2007, 310) define fit statistics as “indices that estimate the extent to which responses show adherence to the modelled expectations.” There are two measures of fit; *infit* and *outfit*. *Infit* statistics indicate the extent to which persons or items fit with the Rasch model, that is, they indicate “the degree of fit of observations to the Rasch-modelled expectations” (Bond & Fox, 2007, p. 310), while *outfit* statistics “are influenced by off-target observations,” and

indicate the extent to which extreme cases and/or unexpected responses are affecting the data (Wright & Stone, 1999, p. 53).

This table summarizes the *infit* and *outfit* statistics for the items used to measure the latent trait.

Table 1: Summary of 54 measured items

	TOTAL SCORE	COUNT	MEASURE	MODEL ERROR	INFIT		OUTFIT	
					MNSQ	ZSTD	MNSQ	ZSTD
MEAN	62.1	105.9	.00	.29	.99	-.1	1.04	.1
S.D.	30.4	.3	1.85	.13	.09	.8	.42	1.1
MAX.	30.4	106.0	3.91	1.01	1.24	1.9	3.22	3.5
MIN.	5.0	105.0	-4.56	.21	.81	-2.4	.18	-2.2
REAL RMSE	.32	TRUE SD	1.82	SEPARATION	5.66	ITEM	RELIABILITY	.97
MODEL RMSE	.32	TRUE SD	1.82	SEPARATION	5.74	ITEM	RELIABILITY	.97
S.E. OF ITEM MEAN = .25								
MINIMUM EXTREME SCORE: 1 ITEM								

Table 3 indicates the extent to which items measured the ability of individuals in the data set. Three items, 34, 35 and 30 exhibited statistically significant misfit ($ZSTD < 2.00$). These items also had large outfit mean square values (> 1.50) which means that their suitability for measurement needs to be assessed.⁹ This table suggests that these items elicited responses from persons that were affected by something other than mathematical competence. These items were not measuring the latent trait effectively. There is, to borrow from the discourse, some ‘noise’ in these items (Linacre, 2002).

These items measure the degree of competence in ordering numbers on an ordinal number line. In each respective language the rater would ask, what translates into English as, “What number is one bigger than 7?” (30), “What number is two smaller than 5?” (34), and “What number is two bigger than 4?” (35).

In our initial review of the English translations of the original German test, we considered these items to be linguistically incongruent with the South African linguistic context. Even the English cohort that we tested struggled with the formulation of these items. The reasons for this may be multiple, but one possible explanation lies in the fact that our English cohort consisted of learners who speak an African language at home, and are schooled in English from Grade 0 through to Grade 12. They receive no instruction in an indigenous African language while at school. However, it appears that they retain certain

9 Linacre (2002) explains that values greater than 1.0 suggest the presence of unmodeled noise or other source of variance in the data. This essentially degrades measurement. On the other hand, values less than 1.0 suggest that the model predicts the data too well. A value of .99 is desirable, since Linacre (2002) argues that MNSQ values between 0.5 and 1.5 are productive for measurement.

linguistic representations of quantity, number and spatial orientation from their mother tongue.

By asking learners from all three linguistic groups in our sample to determine which “number is two bigger than 4,” we were essentially confusing the learners, some of whom interpreted this question as referring to size that can be measured, rather ironically, with a ruler. We effectively confused terminology, signs and concepts for them. Bear in mind that the test was administered orally in clinical interview mode. The children may have been baffled by the issue of whether the word “number,” in the formulation, refers to a digit (that can be signed), a quantity (that can be measured), or a concept (that we have of a number)? Does “two” refer to the digit, the numeral as a sign for measurement or the concept? And does “five” refer to the digit, the quantity or the concept?

Table 2: Item statistics: misfit order

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	INFIT		OUTFIT		PT-MEASURE		EXACT MATCH	
					MNSQ	ZSTD	MNSQ	ZSTD	CORR.	EXP.	OBS%	EXP%
34	10	106	3.13	.34	1.14	.6	3.22	3.5	A-.06	.24	90.6	90.5
35	11	106	3.01	.33	1.05	.3	2.16	2.3	B.13	.25	89.6	89.6
30	25	106	1.93	.24	1.24	1.9	1.59	2.4	C.05	.34	73.6	77.6
31	22	106	2.11	.25	1.03	.3	1.42	1.6	D.24	.32	79.2	79.9
3	82	106	-.90	.25	1.11	.8	1.41	1.8	E.20	.36	79.2	79.1
32	20	105	2.20	.26	1.09	.6	1.40	1.5	F.17	.31	81.0	81.1
47	5	106	3.91	.47	1.08	.3	1.39	.8	G.05	.18	95.3	95.3
17	77	106	-.60	.24	1.16	1.4	1.34	1.8	H.20	.37	70.8	75.8
33	13	106	2.81	.31	1.10	.5	1.33	1.0	I.13	.27	87.7	87.7
7	79	106	-.72	.24	1.20	1.6	1.26	1.3	J.17	.37	71.7	77.1
28	84	106	-1.02	.26	1.02	.2	1.23	1.0	K.32	.35	79.2	80.6
2	78	106	-.66	.24	1.02	.2	1.20	1.1	L.31	.37	77.4	76.4
10	82	106	-.90	.25	1.08	.7	1.17	.8	M.26	.36	79.2	79.1
6	79	106	-.72	.24	1.14	1.1	1.17	.9	N.24	.37	71.7	77.1
MEAN	62.9	105.9	-.11	.32	.99	-.1	1.04	.1			81.1	80.3
S.D.	30.7	.3	1.99	.24	.09	.8	.42	1.1			8.1	8.6

The misfitting nature of items 34, 35 and 30, respectively, suggests that these items, or rather the translations of the original German formulation into English, Sesotho and isiZulu, are having a negative effect on the ability of these items to measure a person’s ability to place numbers on the ordinal number line, which is a component of mathematical competence.

Table 4 indicates the extent to which persons measured responded consistently with items used to determine their mathematical competence. It shows that six respondents out of 106 who were tested provided responses to items that were not anticipated or predicted. In most cases respondents with low mathematical competence answered difficult items correctly when the Rasch model expected them

to answer incorrectly. Since the MARKO-D only requires correct or incorrect answers, it is reasonable to assume that these respondents guessed correctly, or provided incorrect responses to comparatively easy items.

Table 3: Person statistics: misfit order

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	INFIT		OUTFIT		PT-MEASURE		EXACT MATCH	
					MNSQ	ZSTD	MNSQ	ZSTD	CORR.	EXP.	OBS%	EXP%
42	13	55	-1.93	.39	1.26	1.2	6.14	3.7	A .35	.53	77.8	83.1
89	29	55	.04	.35	1.49	2.6	3.20	4.0	B .34	.60	68.5	77.6
73	33	55	.53	.36	1.65	2.9	2.69	3.1	C .32	.60	72.2	80.1
84	30	55	.16	.35	1.46	2.4	2.66	3.2	D .38	.60	66.7	78.2
106	33	55	.53	.36	1.52	2.4	2.50	2.8	E .38	.60	68.5	80.1
50	21	55	-.89	.34	1.31	2.1	2.46	2.5	F .41	.57	63.0	74.7
36	25	55	-.43	.34	1.16	1.1	1.81	1.8	G .50	.59	68.5	75.6
MEAN	32.6	55.0	.54	.37	.99	-.1	1.04	.1			81.1	80.3
S.D.	7.2	.2	.94	.02	.24	1.2	.71	1.1			6.6	3.0

Andrich, Marais & Humphrey (2012, p. 417) consider “random guessing to be the function of the difficulty of an item relative to the proficiency of a person,” and not “a property of the item per se” (418). When a person with a relatively low manifest ability on the latent trait gets a difficult answer right, it is reasonable to assume that the person guessed correctly, unless the person suddenly develops an increased ability on the trait being tested. While other explanations abound, the unexpected response recorded when a person with a low ability endorses a difficult answer suggests that something has gone wrong in the measurement of the latent trait.

Table 4 shows that the responses of six persons out of the 106 tested did not fit with the model’s expectations of how they should have fared. A closer look at their recorded responses suggests that five of the six persons got an item, or a number of items, right that they should have got wrong, while one person “106” got comparatively easy items wrong. A lack of concentration could have affected the answers given by this person.

The question we are faced with is: “now that we know which items and persons are misfitting, what is to be done with them?” The most tempting solution would be to remove faulty items and aberrant responses from our data to get a better fit, but the problem with this is that, in the process, useful information about our items and the persons measured is ignored. It would be more useful to use the information gained from the misfitting items to produce better formulations of the items to better measure the latent trait. However, when it comes to dealing with person fit it would be a good idea to conduct an analysis of item invariance across the three different cohorts to determine whether responses to the items were influenced by something

other than item difficulty and person ability. A differential item functioning (DIF) analysis and an analysis of invariance were conducted to determine whether there were significant differences to report.

The differential item functioning analysis shows that the differences between the groups, which appear to be slight, may indicate that the way in which particular items were translated, scored or communicated to the learners tested could have affected responses to a limited set of specific items. This might suggest that something other than item difficulty and person ability was having an effect on the way in which learners responded to particular items. However, an analysis item polarity did not indicate a definite difference between the three groups for the test as a whole, suggesting that the principle of *unidimensionality* was not violated.

Generally, it appears that the groups had similar responses to the items in the test. With significant differences occurring at items, 11, 42, 43 and 45. These differences ultimately called for an analysis of invariance across the three groups. The analysis indicated that there is little difference between the Sesotho and isiZulu cohorts, but that there might be a case for arguing there is variance between the English and Sesotho and the English and isiZulu responses to the items of the test.

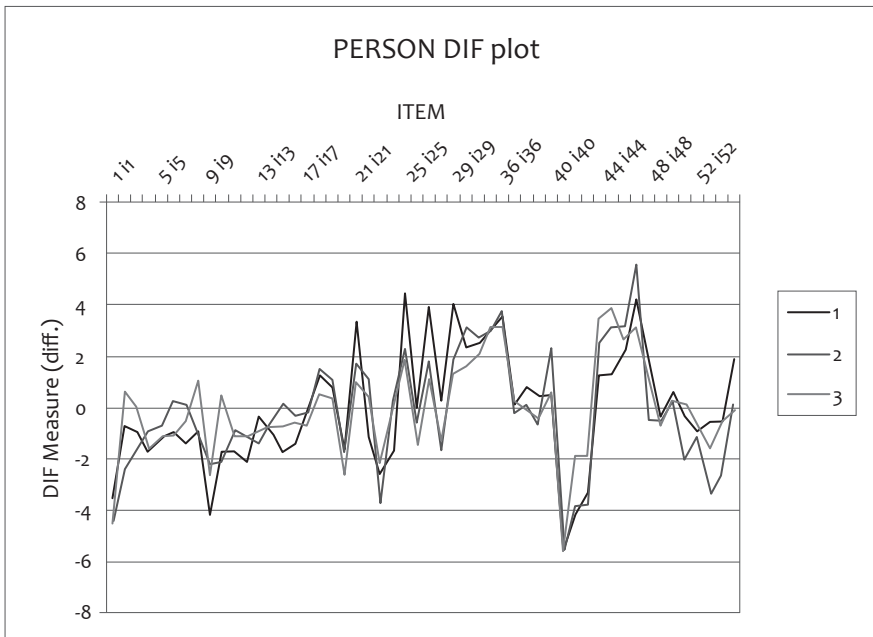


Figure 3: Comparing the performance of individuals in the three language groups

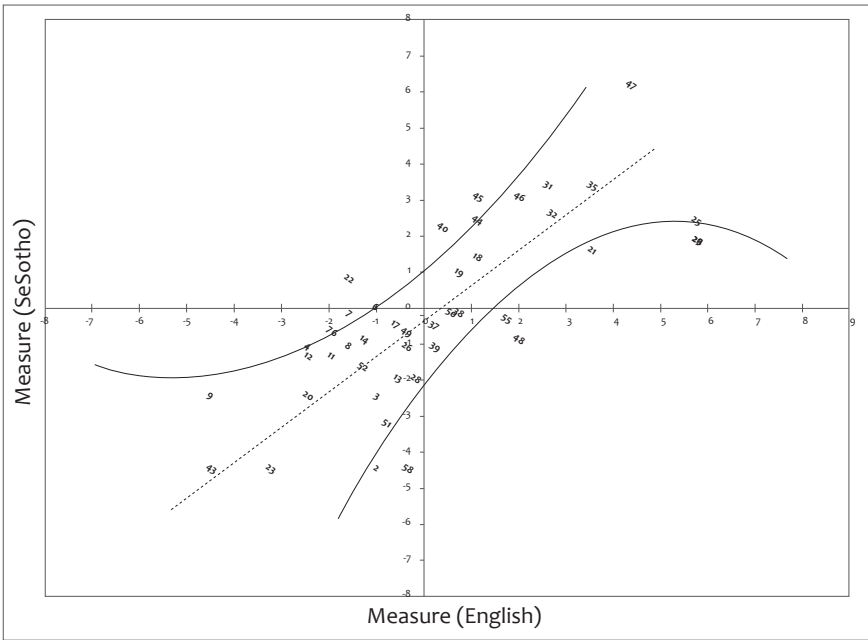


Figure 4: English-Sesotho comparison

Figure 3 indicates the extent to which the English and Sesotho cohorts varied in their responses to the items used to measure the latent trait.

This scatter plot indicates that a notable portion of items (10 of 55) fall outside of the 95 % confidence interval, which indicates that these two language groups differed in the way they responded to these particular items. Ideally all of the items should fall within the confidence intervals to negate the possibility of unmodelled noise affecting person responses to the items. When contrasted with the isiZulu-Sesotho comparison (Figure 3) it is clear that there is more variation in the English-Sesotho comparison than in the former (8 fall outside of the confidence intervals, which still suggests that an extraneous variable is affecting responses to these items, however slight it may be). The same can be said for the English-isiZulu comparison (where as many as 14 items fall outside of the confidence intervals). Notably, when we first assessed the isiZulu translation we considered it to be the most cumbersome and problematic (Welch et al, 2010).

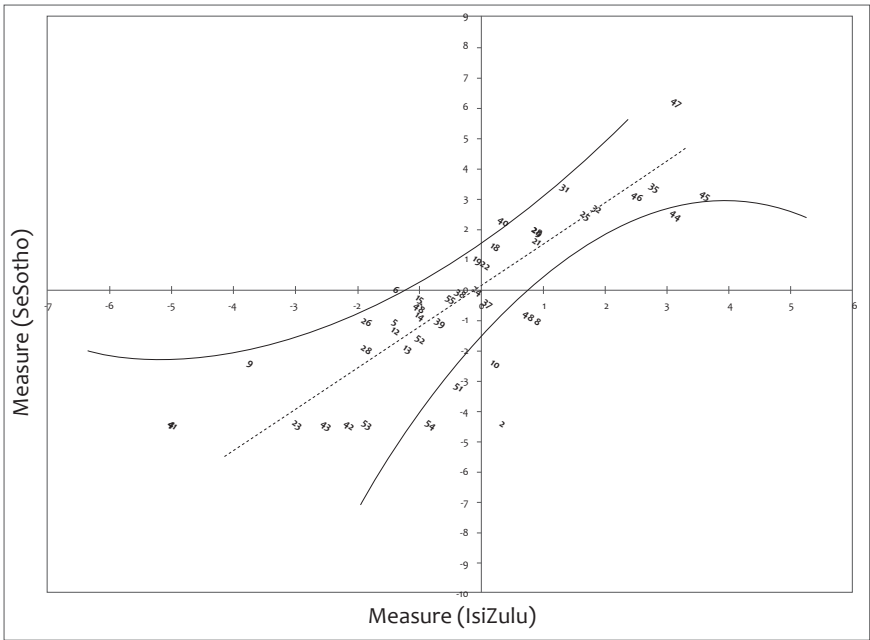


Figure 5: isiZulu-Sesotho comparison

Conclusion

More data are required to make a conclusive assessment of whether or not there are significant differences between the three linguistic groups and whether an added dimension is likely to affect the accuracy of the *MARKO-D*'s ability to measure mathematical competence in this South African setting. However, it was shown that the double translation did have an influence on how particular items were understood by the learners, which affected access to the construct these items were designed to assess.

While it may not be empirically or scientifically sound to make a hard and fast generalization at this point, it does appear that the *MARKO-D* measures basic mathematical competence across various language groups, and that language does not interfere to a significant extent with its ability to capture mathematical competence in learners from three different language groups, for which the test was not initially designed. Even though the linguistic formulation of certain items can be strengthened and rendered more linguistically-culturally sensitive, as it has been in subsequent testing of the first and second cohorts, the test goes a long way to measuring mathematical competence in Sesotho, isiZulu and English learners in a shared urban South African setting.

The *MARKO-D* was designed with language, and the role it plays in conceptual development, in mind. The levels embedded within this instrument are a synthesis

of theories that try to account for the development of mathematical competence. Language plays a big part in this process of development, since the cultural construction of mathematical representations, as in the case of integers, rests and relies on language as a semiotic mediator.

The pupils tested in this study are schooled either in their home language or in what is set to become the language of teaching and learning, i.e. English, yet performance across the cohorts in this test is consistent. In contrast with conventional wisdom, this suggests that the medium of instruction does not play a significant part in the development of early mathematical representations. Why then do we persist with implementing a language policy that suspends our children in a state of linguistic liminality, “betwixt and between” political ideals and global imperatives (Henning & Dampier, 2012; Dampier, 2012)?

Acknowledgements

The research is conducted with funding from the National Research Foundation in a competitive rated research grant for Prof E. Henning and the authors are members of her research team. The work is also supported by the DHET/EU grant programme for research in foundation phase teacher education.

We wish to thank Elbie Henning for her expert assistance in the writing of the article. Deon De Bruin and Annemarie Fritz-Stratmann provided valuable insights and suggestions to earlier iterations of this article, which would not have been finished without their critical input.

References

- Andrich, D (1988) *Rasch Models for Measurement*. London: Sage Publications.
- Andrich, D., & Hagquist, C. (2012). Real and artificial differential item functioning. *Journal of Educational and Behavioural Statistics*, 37(3), 387-416. doi: 10.3102/1076998611411913
- Andrich, D., Marias, I., & Humphrey, S. (2011) Using a theorem by Andersen and the dichotomous Rasch model to assess the presence of random guessing in multiple choice items. *Journal of Educational and Behavioural Statistics*, 37(3), 417-442. doi: 10.3102/1076998611411914
- Benjamin, W (trans. Rendall, S). (1997) The translator’s task. *Traduction, Terminologi, Redaction* 10(2), 151-165.
- Bond, T. G. & Fox, C. M. (2007) *Applying the Rasch Model: Fundamental Measurement in the Human Sciences* (2nd edition). London: Routledge.
- Butterworth, B. (2005) Foundational numerical capacities and the origins of dyscalculia. *Trends in Cognitive Sciences* 14(12), 534-541.
- Carey, S (1985) *Conceptual Change in Childhood*. Cambridge, MA: MIT Press.

- Carey, S., & Spelke, E. (1994) Domain-specific knowledge and conceptual change. In L. A. Hirschfield & S. A. Gelman (eds) (1994) *Mapping the Mind: Domain Specificity in Cognition and Culture* (pp. 169-200). Cambridge University Press: New York.
- Carey, S., & Spelke, E. (1996) Science and core knowledge. *Philosophy of Science*, 63(4), 515-533. Retrieved from <http://www.jstor.org/stable/188065>
- Carey, S. (2001). Cognitive foundations of arithmetic: Evolution and ontogenesis. *Mind and Language*, 16, 37–55. Carey, S. (2004). Bootstrapping and the origin of concepts. *Daedalus*, 133, 59–68.
- Carey, S. (2004) Bootstrapping and the origin of concepts. *Daedalus* 122(1), 59-68.
- Carey, S. (2009) *The origin of concepts*. London: Oxford University Press
- Chomsky, N. (1980) *Studies on Semantics in Generative Grammar*. The Hauge: Mouton Publishers.
- Cook, L. L., Dorans, N. J., & Eignor, D. R. (1988) An assessment of the dimensionality of three SAT-verbal test editions. *Journal of Educational and Behavioural Statistics*, 13(1), 19-43. doi: 10.3102/10769986013001019
- Dampier, G. A. (2012) The (f)utility of additive bilingualism in South African education. *Education as Change: Journal of Curriculum Research* 16(1), 69-81
- Dehaene, S. (2010) *The Number Sense: How the Mind Creates Mathematics* (Revised and expanded edition). New York: Oxford University Press
- Dehaene, S. & Brannon, E. (2011) *Space, Time and Number in the Brain: Searching for the Foundations of Mathematical Thought*. London: Academic Press.
- de la Torre, J., & Patz, R. J. (2005) Making the most of what we have: A practical application of multidimensional item response theory in test scoring. *Journal of Educational and Behavioural Statistics*, 30(3), 295-311. doi: 10.3102/10769986030003295
- DeMars, C. (2010) *Item Response Theory*. Cape Town: Oxford University Press
- De Saussure, F (1986) *Course in General Linguistics*. London: Open Court
- Evans N & Levinson S C 2009. The myth of language universals: Language diversity and its importance for cognitive science. *Behavioural and Brain Sciences*, 32(5): 429-492
- Feigenson, L., Carey, S., & Spelke, E. (2002) Infants' Discrimination of Number vs. Continuous Extent. *Cognitive Psychology* 44, 33-66.
- Feigenson, L., Dehaene, S., & Spelker, E. (2004) Core systems of number. *Trends in Cognitive Science* 8(7) 307-314.
- Gelman, R. & Gallitsel, C. R. (1978). *The child's understanding of number*. Cambridge, MA: Harvard University Press.
- Gelman, R. (1991) Epigenetic foundations of knowledge structures: initial and transcendent cognitions. In S. Carey & R. Gelman (Eds.) *Epigenesis of mind: Essays on biology and cognition* (pp. 293-322). Hillsdale, NJ: Erlbaum.
- Gelman, R., & Butterworth, B. (2005) Number and language: How are they related? *Trends in Cognitive Science* 9(1), 6-10.

- Gerlach, M., Fritz, A., & Leutner, D. (2012). MARKO-T. Mathematik und Rechnen – Training zur Förderung von Konzepten im Vorschul- und frühen Grundschulalter [MARKO-T. A math and arithmetic training on concepts in pre-school and primary school age]. Göttingen: Hogrefe.
- Gopnik, A., & Meltzoff, (1997) *Word, Thoughts, and Theories*. Cambridge, MA: MIT Press.
- 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(3), 70-79
- Henning, E., & Dampier, G. A. (2012) Linguistic liminality in the early years of school: Urban South African children “betwixt and between” languages of learning. *South African Journal of Childhood Education* 2(1), 100-119
- Holloway, I. D., & Ansari, D. (2009) Mapping numerical magnitudes onto symbols: The numerical distance effect and individual differences in children’s mathematics achievement. *Journal of Experimental Child Psychology* 103, 17-29.
- Jackobson, R. (1959) “On Linguistic Aspects of Translation.” In Venuit, L. (2004) *The Translation Studies Reader*, pp 139-147. London: Routledge.
- Le Corre, M. & Carey, S. (2007) One, two, three, four, nothing more: An investigation of the conceptual sources of the verbal counting principles. *Cognition* 105, 395-438.
- Le Corre, M., van de Walle, G., Brannon, E.M. & Carey, S. (2006). Re-visiting the competence/performance debate in the acquisition of the counting principles. *Cognitive Psychology*, 52, 130-169.
- Leslie, A. M. (1994) ToMM, ToBy, and Agency: Core architecture and domain specificity. In L. Hirschfeld and S. Gelman (Eds.) *Mapping the Mind: Domain specificity in cognition and culture* (pp. 119-148). Cambridge, UK: Cambridge University Press
- Levinson S C 2003. Language and mind: Let’s get the issues straight! In: S D Blum (ed.), *Making sense of language: Readings in culture and communication*: 95-104. Oxford: Oxford University Press
- Linacre, J.M. (2002) What do Infit and Outfit, Mean-square and Standardized mean? *Rasch Measurement Transactions*, 16 (2), 878.
- Nazzi, T & Gopnik, A (2001) Linguistic and cognitive abilities in infancy: when does language become a tool for categorization? *Cognition*, 80, 11-20.
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Nielsen & Lydiche.
- Ricken, G., Fritz, A., & Balzer, L. (2012). MARKO - D - Mathematik und Rechnen - Test zur Erfassung von Konzepten im Vorschulalter [MARKO-D – math and arithmetic – test for assessing concepts at pre-school age]. Göttingen: Hogrefe.
- Rips, L. J., Bloomfield, A., & Asmuth, J. (2008) From numerical concepts to concepts of number. *Behavioural and Brain Sciences* 31, 623-642

- Rips, L. J., Asmuth, J., & Bloomfield, A. (2006) Giving the boot to the bootstrap: How not to learn the natural numbers. *Cognition* 101(3), 51-60
- Spelke, E. (2000) Core Knowledge. *American Psychologist*, 55, 1233-1243.
- Spelke, E. (2003) What makes us smart? Core knowledge and natural language. In D. Gentner & S. Goldin-Meadow (Eds.) *Language in mind: Advances in the study of language and thought* (pp. 277-311). Cambridge, MA: MIT Press.
- Starkey, P., Spelke, E. S., & Gelman, R. (1990) Numerical abstraction by human infants. *Cognition* 36, 97-127.
- Vygotsky, L. (1978) *Mind in Society. The development of higher psychological processes*. M. Cole, V. John Steiner, S. Scribner & E. Soubberman (eds and translators). Cambridge, MA: Harvard University Press.
- Vygotsky, L. (1992) *Thought and language*. (edited and revised by A. Kozulin, 6th Edition). Cambridge, MA: MIT Press.
- Wainer, H., Morgan, A., & Gustafsson, J. (1980) A review of estimation procedures for the Rasch model with an eye toward longish tests. *Journal of Educational and Behavioural Statistics*, 5(1), 35-64. doi: 10.3102/10769986005001035
- Welch, N., Dampier, G. A., and Mawila, D. (2011, October) Funda UJabule: Pilot Findings. In E. Henning (Chair) *CEPR Research Seminar*. The seminar was conducted at the University of Johannesburg, Bunting Road Campus.
- Whorf, B. (1956), John B. Carroll (ed.), ed., *Language, Thought, and Reality: Selected Writings of Benjamin Lee Whorf*, MIT Press
- Wright, B. D. (Spring, 1988) Where do dimensions come from? *Popular Measurement*, 1(1) 32.
- Wright, B.D. & Linacre, J.M. (1994). Reasonable mean-square fit values. *Rasch Measurement Transactions*, 8 (3), 370.
- Wright, B.D. & Stone, M.H. (1979). *Best Test Design*. Chicago: MESA Press.
- Wright, B. D., & Stone, M. H. (1999) *Measurement Essentials* (2nd Edition). Wilmington, Delaware: Wide Range, inc.