Is the order of learning numerals universal? Evidence from eight countries and six languages.

Lee Copping[a], Peter Tymms[b], Gabrijela Aleksic[c], Tiago Bartholo[d], Sarah J Howie[e], Mariane Campelo Koslinski[f], Christine Merrell[g], Maša Vidmar[h] and Helen Wildy[i]

Corresponding author: Maša Vidmar

a Teesside University; School of Social Sciences, Humanities and Law, Middlesbrough, TS1 3BX L.Copping@tees.ac.uk

b Durham University; School of Education, Durham University, Confluence Building, Stockton Road, Durham, DH1 3LE, UK, p.b.tymms@dur.ac.uk

c University of Luxembourg; Maison des Sciences Sociales,11, Porte des Sciences,L-4366 Esch-sur-Alzette,Luxembourg, gabrijela.aleksic@uni.lu

d Federal University of Rio de Janeiro; Post-Graduation Program in Education, Avenida Pasteur, 250, sala 234, CEP 22290-902, Rio de Janeiro, Brazil. tiagobartholo@gmail.com

e Stellenbosch University International; 15 Victoria St, Stellenbosch Central, Stellenbosch, 7600, South Africa Howie, sarahhowie@sun.ac.za

f Federal University of Rio de Janeiro; Post-Graduation Program in Education, Avenida Pasteur, 250, sala 234, CEP 22290-902, Rio de Janeiro, Brazil. mckoslinski@gmail.com

g Durham University; School of Education, Durham University, Deceased

h Pedagogical Institute of Slovenia/Educational Research Institute; Gerbičeva 62, 1000 Ljubljana, Slovenia, masa.vidmar@pei.si

i The University of Western Australia, Graduate School of Education; M428, 35 Stirling Highway, Crawley, 6009, Western Australia, helen.wildy@uwa.edu.au

Author Note: The data from England and Scotland was published previously in Cramman et al., (2018) and is used within this study for comparison against the other six countries.

Declaration of Interests: None

Funding: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Is the order of learning numerals universal? Evidence from eight countries and six languages.

Abstract

AUTHOR (2018) proposed that numerical symbol identification may constitute a universally predictive measure of early mathematical development. While a broad pathway to learning number symbols is unsurprising, lack of systematic variation in acquisition order relative to factors such as teaching, age, country, progression stage, is. This study evidences unidimensionality of measurement of the order of ability clusters of numbers, showing that variations are minor across eight countries and, importantly, six instructional languages. This invariance suggests early symbol identification could represent a universal measurement which could a) instructionally inform teaching and learning of classroom mathematics, b) work predictively as an educational research tool and c) offer a foundation for valid international comparisons of the mathematical development of children. Tentatively, this study suggests numerical symbol identification may be a universal measure to assess mathematical cognition in early years education that is unaffected by language of instruction, gender, time of assessment and country.

Introduction

Modern Hindu-Arabic numeracy owes its legacy to three historical advances: distinct symbols for the first nine digits and zero, counting via base 10 and place-value. This 10-symbol approach evolved in Kerala region of India, with the oldest place-value document using base 10 emerging from the same region (Joseph, 2009). These innovations spread globally to emerge as the established way to represent numbers. Although we know the origins of number symbols, our understanding of how we learn to recognise them is not entirely clear.

Nevertheless, research has suggested that the ability to map number words to their Hindu-Arabic counterparts is related to early informal mathematics or 'Number Sense' (conceptually including abilities such as counting, cardinality, arithmetic. - Malofeeva et al., 2004; National Mathematics Advisory Panel, 2008; Siegler & Booth, 2004). Mathematical development is cumulative, with advanced skills and core concept connections developing from consolidation of early skills (Gersten et al., 2005). Failure to grasp early concepts is a key indicator of later mathematical difficulty.

AUTHOR, (2018) suggested that the development of number identification may have a single developmental pathway and that knowledge of children's position along it can be instructionally informative to help foster and measure progress. They hypothesised and empirically tested the proposition that children learn numbers in stages, learning single digits, teens, double digits, triple digits and so on. Between each stage were measurable jumps in difficulty. This suggests that children consolidate earlier numeral knowledge before gradually understanding and applying place-value to identify larger numbers.

Moreover, they showed that this pathway did not vary by sex, country (England and Scotland) socioeconomic background between the beginning and end of the first year of formal education. This is particularly interesting as it suggests a potentially invariant progress measure by which all children could be assessed regardless of circumstances, and to which future mathematical attainment could be anchored to. Last but not least is the potential for it to be used as an effective tool for meaningful international comparisons. In this extension of the original work, we examine if this same pathway exists across multiple countries and languages.

How do individuals identify numbers?

Several cognitive models of Hindu-Arabic number recognition, based on the widely studied patterns of functionally dissociated deficits in neurological patients, have been proposed (Dotan & Friedmann, 2018; McCloskey et al., 1985). While reviewing all of the technical differences between these models is beyond the scope of this paper, they propose the use of syntactic frames in number identification, which then transform into number word frames once digit size and the position of 0 is accounted for. Briefly, when a numeral is presented, lexical processing identifies its elements while syntactical processing assesses relations between elements to arrive at the correct identification. Lexical retrieval requires the selection of the correct lexical class (such as Ones, Teens and Tens) and then the correct number word from within that class. A frame based on the presented numerals is generated from which the number words (*Ones, Tens, Teens*) and decimal words (*Hundreds, Thousands*) from the lexicon can be inserted, as well as combined with function words such as 'and'. For example, 1234 becomes 1 (*1 Ones Thousand*) 2 (*2 Ones hundreds*) 3 (*3 Tens*) and 4 (*4 Ones*). The lexicon of each word is also linked to the phonological and symbolic representations of the number word to allow for verbal or written production of the numerals (in this example*, one thousand, two hundred and thirty-four*). This process is underpinned by a number of independent sub processes, which will not be detailed here but see Dotan and Friedmann (2018) for a comprehensive overview of recent cognitive models of numerical processing.

It is important to recognise that numerals can be comprehended successfully but inaccurately produced verbally or in writing (Benson & Denckla, 1969; McCloskey et al., 1985). There is also acknowledgement that some numbers can be identified by alternative cognitive routes because they are 'lexicalized'. These include common numbers such as 100, or numerals read as dates such as '2001' (Cohen et al., 1994). Finally, most of these processes are functionally independent of word reading (Dotan & Friedmann, 2018).

How, and in what order, do children learn number symbols?

The models above outline how number symbols are processed into an identifiable number output in cognitively functional individuals. What we know less about is how children acquire the knowledge required to functionally operate the syntactic framework and number word frames. It may first appear that learning to identify numerals in order from ‘1’ upwards to ‘n’ should be unsurprising. But why should we expect arbitrary mappings of phonemic to symbolic representations to occur in any order at all? This requires a brief explanation of the context within which learning early numeracy skills occurs. Numeracy develops in overlapping phases between counting and quantity mapping, often referred to as informal mathematics (Purpura & Ganley, 2014) and are the bases upon which formal mathematical skills develop. Number identification however does not conform to the definition on formal or informal mathematics (Baroody & Wilkins, 1999), which has led to the recent suggestions that ‘number knowledge’, which includes identification, may be a bridge between the formal and informal (Purpura et al., 2013), making it a vital prerequisite of advanced skills development.

Although generally emerging at approximately age 3 (Gelman & Gallistel, 1978), symbolic mapping can occur as early as 18 months (Mix, 2009). The use of symbols is complex and requires three abilities: awareness of the relation between symbol and referent; mapping the corresponding elements from one to the other and drawing an inference about one based on knowledge of the other (DeLoache et al., 1999).

Number word meanings develop in conjunction with counting, and understanding magnitude and cardinality (Carey, 2004; Krajewski & Schneider, 2009; Sarama & Clements, 2009), although use of count words is often independent of actual counting in the very young (Wagner & Walters, 1982). Litkowski et al., (2020) found that by age three, 55.8% can identify ‘1’ while 3.8% identify ‘15’. By age five, these proportions are 95.7% and 40.9% respectively, suggesting rapid progress with age. These developments coincide with neurological changes suggestive of increasing specialisation in the left intraparietal sulcus (Merkley & Ansari, 2016) as well as alongside growth in other informal skills (particularly counting, cardinality and one-to-one correspondence (Litkowski et al., 2020). Children begin to understand that some single digits such as 1 to 5, represent smaller quantities than others such as 6 to 9, sometimes independently of precise meaning and magnitude quantification (Le Corre & Carey, 2007). As children map words to symbols, proficiency at this grows quickly, potentially supported by anchoring to earlier mappings (Mix et al., 2014). Indeed, Baroody et al., (1983) demonstrated that reading and writing of number symbols in a stepwise manner as they progress through early schooling.

Multi-digit numbers can also be identified from as early as three-and-a-half (Mix et al., 2014), suggesting that number identification is possible without formal instruction, although generally this is a more difficult skill (Claessens et al., 2014). A clear understanding of multi-digit identification is complex as it includes: some understanding of place-value, associating multi-unit names with written position, some comprehension of the distinctive nature of zero, the link between written position and base-10 values, that unlike in reading, size increases right to left, as well as linguistic irregularities in some languages, such as -teen and -ty in English (Baroody et al., 1983; Byrge et al., 2014; Fuson, 1990). These processes would culminate in the number word frames discussed earlier (Dotan & Friedmann, 2018). Mix et al., (2014) demonstrated that pre-schoolers have some informal understanding of place-value, which allows some comprehension of multi-digit numbers and that performance increases with age. Ceiling effects begin to emerge in second graders, presumably coinciding with formal place-value instruction. Mix et al. further suggested that this learning occurs through exposure to appropriate but numerically complex stimuli,from building blocks to phone numbers, in the developmental environment. AUTHOR, (2018) also suggested (based on Benford's Law, 1938) that the distribution of numerals in natural number sets (lower numerals being more frequent) may also play a role in consolidating the order in which numerals are learned. Common benchmark numbers such as 10, 100 and higher appear often in informal learning games, parental interactions and early instruction (Byrge et al., 2014), which may explain why some large multi-digit values are sometimes more easily identified by young children than expected (Byrge et al., 2014; AUTHOR, 2018) and may suggest some form of 'lexicalised' response (Cohen et al., 1994).

Taken together, evidence on early symbol identification suggests that number identification is progressive from simple single digits to more complex multi-digits. AUTHOR, (2018) empirically tested this pathway on large representative samples of children in the UK and demonstrated that this indeed appeared to be the case, with numerals 1 to 5 being learned first, followed by 6 – 9, teens, two digits and three digits. There were noticeable jumps in difficulty between each stage, the size of which decreased in line with increased progress. Some numbers (e.g., 100) were, as suggested above, more easily recognisable relative to other numbers of the same size but, overall, a single pathway with stepwise difficulty jumps (Baroody et al., 1983) was clearly identifiable.

The ability jumps are important to note. Why should there be a noticeable jump between 5 and 6? It is, after all, one extra symbol? Assuming we just learn symbols in order would not predict this to happen, yet this and other jumps like it are evident in the data. This pathway suggests that there is some form of cognitive shift at various points in the sequence. If these pathways represent changes in cognition, then the question becomes, are these too universal? If so, then knowing them can be useful to early years researchers and practitioners alike in assessing early competencies in young children and delays in this progression may account for variation in later mathematical ability.

The predictive power of number identification ability

Research consistently demonstrates that symbolic number processing is a key predictor of later mathematics development and achievement. Moreover, studies have consistently found a greater correlation between early symbolic measures of number representation and mathematical achievement when compared to non-symbolic measures (Lyons et al., 2014; Martin et al., 2014). Martin et al., (2014) suggest that symbolic number identification and symbolic number comparison were better predictors of mathematical achievement in Grade 1 (fluency, computation and math problem solving) when compared to counting (procedural and conceptual) measures. Lyons et al., (2014) also found that symbolic number processing predicted arithmetic ability more strongly than non-symbolic processing.

Further studies report that a child´s early ability to identify numbers is a good predictor of later attainment at the beginning of elementary education and later in their school career. Chard et al., (2005), observed a correlation of .58 between number identification and number knowledge tests at the beginning and end of the academic year for both kindergarten and first grade. Jordan et al., (2007) found that early number competency (number identification, quantity discrimination, identifying missing number in sequences) at the end of kindergarten is a good predictor of mathematics outcomes in first grade. A similar tendency was found at the third grade, when mathematics becomes more complex and one could expect a weaker correlation. Jordan et al., (2009) found a correlation between early number competence and later math achievement measures at the end of 3rd grade that varied from .60 to .52. Furthermore, this study also observed that the rate of growth in early number competences was a good predictor of mathematics performance in 3rd grade. AUTHOR (1999) found a substantial correlation between number identification for children on entry to school in the UK, age 4 years, and mathematics (.60) at age 7. The correlation was similar for reading (.64) at age 7. Predictive validity of number identification measures consistently suggests that it is a strong, possibly the strongest, predictor of later formal mathematics abilities (Lembke & Foegen, 2009; Purpura et al., 2013)

Finally, lack of fluency in identification of numbers, among other measures, can be used as a reliable indicator of potential mathematics difficulties (Gersten et al., 2005). These findings all point to the potential of number identification measures as a diagnostic and screening instrument and for early intervention for children starting school.

International variation and number identification

Although there are distinct advantages to number identification being a useful predictive classroom and research tool, it is reasonable to suspect that this ability varies internationally due to substantive differences in language and cultural practices. We now discuss each in turn. Language skills and number identification are linked. Identifying number symbols correlates with early vocabulary and phonics (LeFevre et al., 2010) while early difficulties in numeracy often co-occur with language difficulties (Purpura & Reid, 2016). This is even the case though the relation between print and speech in symbolic numeral representation is very different to that of text (Mix et al., 2002). Language differences do perhaps impact symbol interpretation, however. Seron and Fayol (1994) showed that Belgian Walloon speakers (where multi-digit decades are regular) were more accurate in transcoding tasks than French speakers where some decades are irregular e.g. 70 is *soixante-dix* (*sixty-ten*) and 99 is *quatre-vingt-dix-neuf* (*four-twenty-ten-nine*). In a novel design to better control for culture, data showed that there are more errors on inverted than non-inverted systems, for example, where the number word is *one-and-twenty* rather than *twenty-one*. Pixner et al., (2011) showed that in Czech children, whose language is unique in that it has inverted and regular systems held in parallel, errors were greater using the inverted system. Miller et al., (2005) found that the base-10 structure of symbol identification in

Chinese seemed to allow for children to count to 100 earlier than English speaking peers. Language may thus impact on how and in what order children begin to learn to recognise numerals.

Isolating sources of variation between language and culture is difficult, Pixner et al's., (2011) study being a rare exception. There are a great many potential cultural confounds that may affect how children learn early number concepts. Age of formal schooling varies dramatically, ranging from age 5 in England to age 7 in Bhutan (Chartsbin, 2018). Some countries lack compulsory starting ages altogether. Pre-school provision is equally as variable (Oberhuemer, 2005). Related to this is the value placed on children's informal knowledge on entry to preschool or school. In some cases, informal knowledge such as finger counting (necessary for learning base-10) can be viewed as a disadvantage rather than an asset (Dehaene, 2011). Teaching methods also vary from abstract manipulations to concrete approaches (Dehaene, 2011). Representation of numbers in the curriculum can also differ, with more concrete methods being prevalent in Asia. For example, the Chinese regular notation consists of only 13 symbols for the digits 1 through 9 and numbers 10, 100, 1,000, and 10,000. The number 2,342 is written as '2 1000 3 100 4 10 2'; a word-for-word transcription of 'two thousand three hundred forty-two' (Dehaene, 2011). Thus, Chinese children learn numbers by simple rules such as 11 = ten one, 12 = ten two, 20 = two ten,) whereas American children learn different number rules from 0 to 10, from 11 to 19, and tens from 20 to 90. By age 4, Dehaene (2011) reviews evidence that Chinese children could count to 40 whereas American children counted only to 15, taking a year to catch up .

Although evidence suggests differences between Chinese and American children in terms of rate of counting ability, Miller et al., (1995) noted that decade boundaries seemed to be increasingly difficult for both cultures. While Chinese students may accelerate more quickly in counting ability than English speakers, they still appear to share the same difficulties. AUTHOR, (2018) also showed that despite age and curricular differences within the first year of schooling in England and Scotland, the order in which children learned to identify numbers was invariant across the two countries. We take this as tentative evidence to suggest that although children may have different rates of progression in identifying number symbols, the order in which they learn them, and the difficulty jumps they encounter, are potentially universal.

Universal measures

There have been attempts to produce common measures for use in countries across the world, including the Trends in Mathematics and Science Study (TIMSS). Beginning in 1995, its most recently completed iteration (TIMSS, 2015) involves 57 countries. International comparisons are made, using Item Response Theory (IRT) to construct scales for each of Grades 4 and 8 (Mullis et al., 2016) providing rich information about students' mathematics attainment across different languages and cultures. Scaling however, is within-grade and so while TIMSS measurement scales are potentially universal, they report attainment at single time-points within grades.

Mathematics curricula often present content areas such as number sense, operations, measurement, shape and space, data analysis, which may be considered to form different dimensions (Burg, 2008). Multi-dimensions are problematic for psychometric measurement models, which assume unidimensionality. Burg's analysis suggested that mathematical skills overlap and that students do not learn skills in isolation. A mathematics test including questions spanning a range of content areas, could thus theoretically represent a unidimensional construct. We are not currently aware of models that successfully reflect this ideal. Furthermore, Burg proposed the possibility of constructing a vertical scale covering students over a range of ages. Further research in this area (Stenner et al., 2015) led to the development of the Quintile Framework for Mathematics, which proposed a developmental taxonomy of mathematics skills and concepts. The Quintile Framework, in the USA is a vertical scale for mathematics skills and understanding, covering Kindergarten to Grade 12, and has been successfully used in 24 countries (www.metametricsinc.com). It should be noted however that researchers in the past have often found that early mathematics abilities often form multi-dimensional structures (for instance, Ryoo et al., 2012).

Although international studies have progressed towards generating universal measures of mathematics, the methodology proposed by Burg (2008) and Stenner et al., (2015) suggests vertical scaling could assist in equating development across countries. The next challenge is to explore the possibility of generating universal measures using such techniques.

The current study

The reviewed literature and an understanding of the logic behind number systems provide a basis for supposing that simpler numbers will be learned before more complex numbers. This study aims to extend the findings of AUTHOR, (2018) by aiming to replicate the hypothesised pathway from 1 to 999 through learning number identification across different languages and countries. If the findings of this original paper hold, we would expect to see the same pathway demonstrated across different languages, cultures and countries. The scale should also demonstrate progression over the first year of formal education. In short, it should be possible to empirically demonstrate near universal invariance and progress across any demographic classification when it comes to learning number symbols. Should this be the case, then we may have a universal measure, predictive of future attainment, that could be used for international comparisons in the early years. The research hypotheses for this study are thus as follows:

1) A unidimensional scale for numbers 1 through 999 should be identifiable.
2) Identifiable difficulty clusters of numbers should be present in the scale.
3) A scale should be invariant across languages, countries, sex, and educational stage.
4) Children should make progress on this scale over the first year of formal schooling.

Method

Sample

Data from schools was gathered from eight different countries using the International Performance Indicators in Primary School assessment (IPIPS, http://www.ipips.org/home) as part of existing research projects operating in Australia, Brazil, England, New Zealand, Scotland, Serbia, Slovenia and South Africa's Western Cape (AUTHOR, 2019; AUTHOR, 2019; AUTHOR, 2014; 2016; 2017; AUTHOR, 2009). The Australian sample is divided into two complete areas (A and B) and the Western Cape sample was split among schools of different primary language mediums (English, Afrikaans and isiXhosa). In total, we measured eleven areas over eight countries on a combined sample of 18,531 children who completed both the baseline and the end of year assessment. Languages spanned four language groups (Germanic (English, Afrikaans), Romance (Portuguese), South Slavic (Serbian and Slovenian) and Bantu (isiXhosa)). Some notable differences in numerical language exists between these groups beyond pronunciation and unit names, particularly around differences in convention for numbers in the teens.

Mean ages across sampled countries ranged from 4 years, 6 months to 7 years, 1 month years at baseline and 5 years, 4 months to 7 years, 5 months years at follow up. See Supplementary Information for full demographic details and for an accounting of sampling and curriculum differences due to national contexts. As this secondary analysis used data aggregated from multiple projects, ethnicity, socioeconomic and pre-education background were not comparable due to local recording differences. However, AUTHOR, (2018) demonstrated that differential item functioning (DIF) was not evident in such measures across England and Scotland where recording mechanisms were comparable.

Table 1: Summary of data from Countries and Areas

Area	N at Baseline	N at follow up	Age (years & months) at Baseline	Age (years & months) at follow up	Sex (% female)	School Language	Additional language (%)
Australia A	1310	1280	5Y 8M	6Y 4M	46	English	3
Australia B	3457	3475	5Y 5M	6Y 6M	48	English	11
Brazil	162	146	5Y 2M	5Y 10M	54	Portuguese	0
England	6316	4544	4Y 7M	5Y 4M	49	English	17
N. Zealand	2320	1440	5Y 1M	6Y 11M	48	English	14
Scotland	6627	6627	5Y 11M	5Y 10M	50	English	2
Serbia	159	159	6Y 2M	6Y 9M	51	Serbian	28
Slovenia	328	328	6Y 3M	6Y 5M	48	Slovene	2
W. Cape A	1198	644	6Y 11M	7Y 4M	51	English	19
W. Cape B	1451	471	7Y 10M	7Y 5M	46	Afrikaans	5
W. Cape C	999	697	6Y 10M	7Y 5M	51	IsiXhosa	2

Although the samples from England, Scotland and the Western Cape are nationally representative, samples from other locations are not necessarily so (see Appendix 1 for sampling strategy information within each country). However, if the theoretical pathway proposed is universal (as hypothesized) this should not be an issue. The difficulty jumps expected should still emerge regardless of area, language, school type, assessment medium and other sources of difference. In this case, the diversity of the samples should not be problematic for this analysis.

Ethical approval from relevant, university Ethics Committees was granted prior to commencement in all participating countries. Local permissions for data collection and anonymised usage for research were also obtained.

Measures and Procedure

Children's abilities were measured on the first and second occasions using the IPIPS assessment. The assessment includes measures of language,such as letter recognition, phonics, word recognition and reading, and mathematics, such as number reading, counting, cardinality, simple addition, and subtraction. Children are assessed one-to-one by a teacher or researcher for both the baseline assessment and end of year follow-up within their school environment. The assessment can be delivered via an electronic device such as a smart-phone, tablet, or computer, or on paper, and takes approximately 20 minutes to administer.

Using the computer adaptive version, the number identification subscale presents the child with a Hindu-Arabic numeral and is asked "What is this number?" after which a verbal response is recorded as right or wrong. For multidigit numbers, the correct articulation of the word must be produced to be considered correct. For instance,121 is *one hundred and twenty-one* not *one-two-one*. No prompts are provided by the assessor. Numerals are presented in an approximate difficulty order from single digits, teens, double and triple digits. Numerals 0 through 9 are presented in the following order: 4, 1, 3, 2, 5, 7, 6, 9, 8, 0, with the exception being Slovenia(see Table 2). Children then progress though three teens, three two-digits and five three-digits (6 in Slovenia) numerals. A maximum of 21 items are presented (with an exception of 22 in Slovenia). If a child answers four items incorrectly or is unable to answer, or makes three consecutive errors or omissions, the assessment ends. Responses to items after the stopping rule has been activated are recorded as missing rather than incorrect. There is no time limit to the section. Reliability of this section exceeds .90 (AUTHOR, 2018).

The sample is split into two groups for the purpose of this study. Group 1 used the standard computer adaptive section with random numbers. Group 2 used a paper version where numbers were specified in advance for teens, two-digits and three-digits. Scoring mechanisms and stopping rules are consistent across methods. Table 2 shows the assessment composition and numeral order by country. For group 2 countries, the selected numbers to be assessed were locally determined by practitioners and researchers. Equivalency of the method of data analysis was examined via DIF analysis. Only the number 12 showed evidence of DIF, appearing to be slightly easier in the computer version than on paper (-2.43 and -4.37 logits respectively). The two assessment methods are thus considered equivalent for this analysis.

Table 2: Numerals shown in different Countries

Group	Country	Units (0-9)	Teens (10-19)	Tens (20-99)	Hundreds
1	Australia	4 1 3 2 5 7 6 9 8 0	Rnd x 3	Rnd x 3	Rnd x 5
1	England	4 1 3 2 5 7 6 9 8 0	Rnd x 3	Rnd x 3	Rnd x 5
1	N. Zealand	4 1 3 2 5 7 6 9 8 0	Rnd x 3	Rnd x 3	Rnd x 5
1	Scotland	4 1 3 2 5 7 6 9 8 0	Rnd x 3	Rnd x 3	Rnd x 5
2	Brazil	4 1 3 2 5 7 6 9 8 0	14 19	33 92 41 60	136 342 563 742 901
2	Serbia	4 1 3 2 5 7 6 9 8 0	10 15 13	20 40 55 25 43 36	100 281 479
2	Slovenia	4 1 7 9 5 0	11 15 10 12	20 40 55 41 68 92 25 43 36	300 281 479
2	W. Cape	4 1 3 2 5 7 6 9 8 0	10 15 13	20 40 55 25 43 36	100 281 479

*Rnd = Random Number

Groups 1 and 2 had differences in the administered numerals. The data for Group 1 used all numbers from 1 to 9 as well as 0; thereafter random samples of numbers from groups of numbers were presented to the students. Group 2 used specific numbers throughout.

Data Analysis

A Rasch measurement model (Rasch, 1960; Bond & Fox, 2015) was implemented for analysis because it seeks to establish a unidimensional latent construct, the existence of which can be tested. It can also determine if relative difficulty of items varies by groups (DIF), handle missing data and, be used to assess other threats to the validity. Initially, Rasch models were constructed for all items and all students. Reliabilities were recorded and, for items, given the large samples, these were expected to be above .90. For students they were expected to be above .70. Wright maps, which plot item difficulties and students' abilities on the same scale, were compared against theoretical expectations. Item fit was checked using Infit (Information-weighted mean square residual goodness-of-fit statistic) mean-square statistics, which is appropriate due to its greater sensitivity to the inlying patterns of overall performance than outliers and extremes. Items with figures greater than 1.30 or less than 0.70 and significant to $p<.05$ were identified (Smith, 2000). Analyses were conducted using Winsteps (V4.0.1)

Results

In this section, we begin with the descriptive statistics for the measure before tests for each hypothesis are examined in turn. An additional analysis regarding the potential for tautological findings is then conducted.

Tables 3 and 4 show the key statistics for the measurement instrument by country for the items and participants (persons). The statistics were calculated separately for each country. For items, reliabilities are consistently close to 1.00. For persons, reliabilities across countries are in excess of .90 at the start of year and in excess of .80 at the end of year, the exception being New Zealand. Combined with the separation statistics, the results suggest that a consistent hierarchy of item difficulties is identifiable in the data and that individuals of low and high ability can be reliably identified. Mean Infit values are consistently close to 1.00 suggesting that the model shows good overall fit.

Table 3: Descriptive statistics for the items by country

Country	Start of Year			End of Year		
	Reliability	**Separation**	**Mean Infit (SD)**	**Reliability**	**Separation**	**Mean Infit (SD)**
Australia A	0.96	5.22	0.73 (0.30)	0.94	3.90	0.83 (0.39)
Australia B	0.97	6.21	0.83 (0.22)	0.98	6.41	0.83 (0.27)
England	0.97	5.95	0.81 (0.31)	0.97	6.06	0.85 (0.23)
N. Zealand	0.94	3.91	0.77 (0.78)	0.94	3.99	0.77 (0.57)
Scotland	0.98	6.94	0.85 (0.22)	0.99	10.96	0.86 (0.22)
Brazil	0.98	7.51	0.92 (0.20)	0.99	9.07	0.82 (0.40)
Serbia	0.99	13.08	0.82 (0.19)	0.98	6.71	0.86 (0.25)
Slovenia	1.00	19.09	0.89 (0.23)	0.99	10.18	0.94 (0.19)
W. Cape	1.00	65.97	0.77 (0.21)	1.00	22.97	0.93 (0.13)
Overall	0.99	14.04	0.83 (0.16)	1.00	15.68	0.85 (0.16)

Table 4: Descriptive statistics for the persons by country

Country	Start of Year			End of Year		
	Reliability	**Separation**	**Mean Infit (SD)**	**Reliability**	**Separation**	**Mean Infit (SD)**
Australia A	0.91	3.13	0.83 (1.05)	0.85	2.35	0.70 (0.98)
Australia B	0.92	3.29	0.85 (1.00)	0.88	2.73	0.67 (1.00)
England	0.93	3.55	0.87 (0.97)	0.90	3.02	0.77 (1.18)
N. Zealand	0.91	3.21	0.83 (0.95)	0.77	1.84	0.60 (1.32)
Scotland	0.92	3.46	0.82 (1.06)	0.84	2.33	0.69 (1.11)
Brazil	0.90	3.07	0.80 (1.04)	0.92	3.30	0.69 (1.38)
Serbia	0.97	5.76	0.79 (0.85)	0.93	3.64	0.78 (0.62)
Slovenia	0.96	5.09	0.83 (0.85)	0.92	3.34	0.82 (0.76)
W. Cape	0.97	6.19	0.71 (1.06)	0.81	2.07	0.81 (1.26)
Overall	0.93	3.77	0.85 (0.99)	0.88	2.66	0.72 (1.12)

Unidimensionality

The first test is to examine if the items within the assessment form the hypothesised unidimensional pathway. Table 5 illustrates the results from a principal components analysis, calculated separately for each country, with the amount of variance explained by the measure, the first and the second contrast. As evident in Table 5, overall, the measures explain 72.8%

of the variance at the start of the year and 78.5% of the variance at the end of the year. First and second contrasts explain very little additional variance, the highest being Brazil at the start of year and Serbia at the end of year. This analysis suggests that across the nations studied and when combined, the number identification scale is unidimensional.

Table 5: Results of PCA Analyses for each country

Country	Start of Year Percentage of variance explained by…			End of Year Percentage of variance explained by…		
	Measures	1st Contrast	2nd Contrast	Measures	1st Contrast	2nd Contrast
Australia A	76.7	0.5	0.5	78.7	0.4	0.4
Australia B	74.4	0.5	0.5	79.3	0.4	0.4
England	70.5	0.8	0.7	79.6	0.4	0.3
New Zealand	73.5	0.7	0.6	84.1	0.5	0.4
Scotland	75.7	0.4	0.4	80.8	0.3	0.3
Brazil	71.0	5.0	3.3	81.8	2.9	2.7
Serbia	74.3	2.3	2.2	64.8	4.6	4.4
Slovenia	70.1	2.9	2.3	68.8	2.8	2.6
W. Cape	72.5	2.1	2.0	69.0	3.0	2.1
Overall	72.8	0.5	0.5	78.5	0.4	0.4

Are there identifiable ability bands?

Wright maps were used to examine if the difficulties of various items clustered in the order hypothesised and demonstrated in the previous works of AUTHOR, (2018) can be replicated. Figure 1 shows the Wright map for all items and persons combined on the same scale. This plots the difficulty of the items (right) against the ability of the children (left) on a single logit scale (far left). The numeral bands identified by AUTHOR, are clearly visible although the interface between teens and two-digit numbers is narrow. 1 to 5 are approximately between -14 and -12 logits, 6 to 9 between -10 and -7, teens between -5 and -2, two digits between -2 and 2 and three digits largely greater than 6 logits. Some numbers (100 and 300 for example) also appear easier than expected. One will also note that individuals (on the left) extend the entire range of the scale. Figure 2 illustrates the item bands more clearly via a box and whisker plot, demonstrating where some items are slightly

```
MEASURE     PERSON - MAP - ITEM
                <more>|<rare>

   10        .######  +
                   .  |
    9             .# T+
                   .  |T 742
    8             .#  +  563      901
                   .  |  Random 3 digit number 1, 2, 3, 4 and 5
    7             .#  +  281      342      479
                   .  |  136
    6              .  +
                      |
    5              .  +
                  .#  |S
    4 .############   +
                   . S|  300
    3              .  +
                   .  |
    2             .#  +
                .###  |  31 37 51 57 71
                  .#  +  21 32 34 35 38 39 49 52 53 54 58 59 60 61 68 73 75 79 81 87 89 91 92
95 96 97 98
                  .#  |  23 24 26 27 28 29 30 36 41 42 43 45 46 47 48 50 55 56 63 64 65 67 69
70 72 74 76 77
                         78 80 82 83 84 86 88 90 93 94 74 76 77 78 80 82 83 84 86 88 90 93 94

    0            .##  +M 25       33       62       66       85
                  .#  |  40       44
   -1         .#####  +  22
                 .##  |  100       13
   -2            .## M+  15       19       99
                  .#  |  12       20
   -3            .##  +  16       17       18
                  .#  |  14
   -4             .#  +
                 .##  |S
   -5           .###  +  11
                 .##  |
   -6             .#  +
                  .#  |
   -7           .###  +  10
                  .# S|  9
   -8           .###  +  8
                   .  |T 0
   -9             .#  +  6
                  .#  |  7
  -10              .  +
                 .##  |
  -11                 +
                  .#  |
  -12              .  +  5
                  .#  |  2
  -13              .  +  3        4
                  .# T|
  -14              .  +  1
                   .  |
  -15           .###  +
                <less>|<freq>

 EACH "#" IS 469: EACH "." IS 1 TO 468
```

Figure 1: Wright map demonstrating item and person difficulties

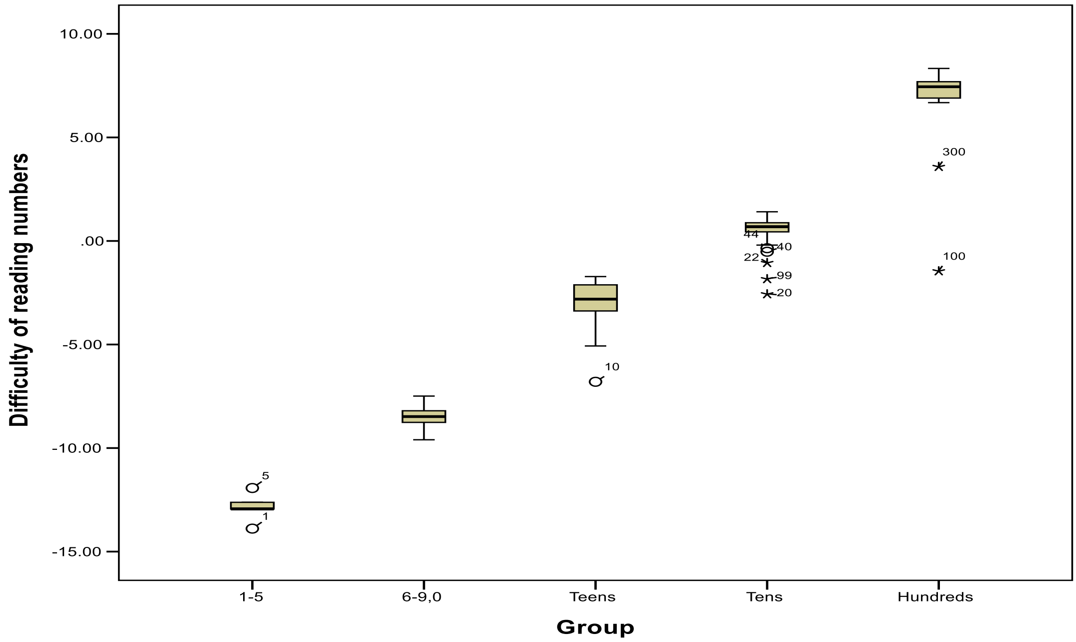
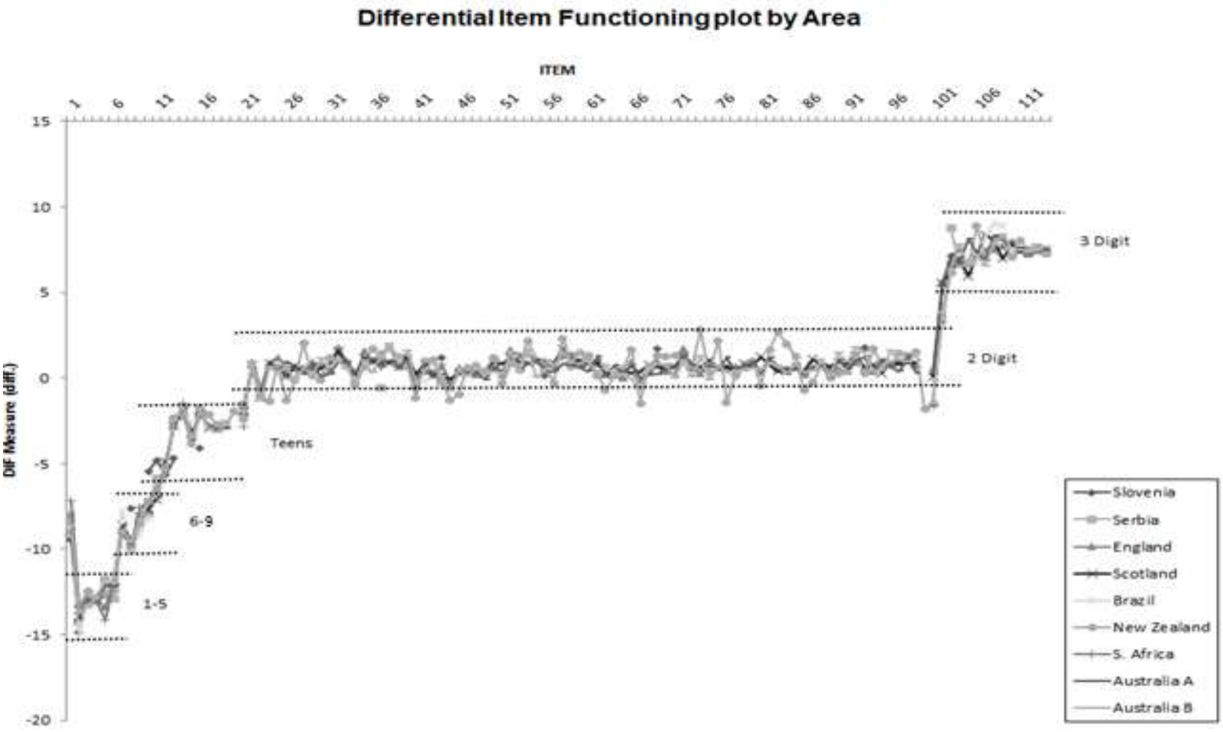


Figure 2: Box and Whisker plot of items by their hypothesised band.



*Note that 0 is the first item on the chart but is more difficult than items 1-5 (appearing in band 6-9). Dotted lines are inserted to aid interpretation by showing approximate positions of clusters (and do not represent confidence intervals).

Figure 3: Item difficulties by country

easier or harder than anticipated while still being in the general difficulty area. Figure 3 shows the pattern for individual countries from a Differential Item Functioning (DIF) analyses of all the data together. One will note the consistency of item difficulties within each band across nations. This analysis suggests that the numeral bands 1-5, 6-9, teens, two-digits and 3-digits are empirically consistent and observable.

Invariance

DIF analysis was further used to examine if items differed substantively and significantly in their difficulty level between the identified groups of interest. These were sex, country, language, additional language and occasion. In group 1, as all 4 nations were native English speakers, language was not considered but the presence of an additional language was. The presence of DIF was examined using Mantel-Haenszel statistics and deemed a potential issue in items where there was both a) substantive differences in item difficulty between groups (defined as >0.64 logits) and b) the Mantel-Haenszel statistic was significant at $p < .05$ (Linacre, 2015).

For sex and additional language, there was no evidence of DIF across any of the items.

For DIF by country, 5.7% comparisons were recorded with $p<.05$ and DIF>0.64. The largest difference was of 2.5 logits $p=0.0011$ for Item number 73, which was easier in New Zealand than in England. Generally, DIF contrasts were small (≤ 2.5 logits) in comparison of the range under consideration (22 logits).

Within most countries there was a single dominant language. However, in South Africa's Western Cape, the assessments were carried out in three languages, English, IsiXhosa and Afrikaans, with approximately equal numbers of students. Accordingly, DIF was examined only for IsiXhosa and Afrikaans against the other languages. This involved

130 comparisons and, of these, 30 (23%) were significant (*p*<.05) and exhibited DIF>.64. The largest contrast was 2.50 for number 281, which was easier in IsiXhosa than Serbian. The presence of DIF for items by language, over and above country DIF, is at a higher rate than would be expected under the null hypothesis (5%). However, as with country DIF, the contrasts were generally small (<2.5) in comparison of the range under consideration (22 logits).

In terms of testing occasion, between the start and the end of the year, there was one instance of DIF for the number 4, which was easier at the start of the year than the end by 1.06 logits. All other item difficulties remained relatively constant between the start and end of the first year at school.

Progress in the first school year

The mean ability of each sample increases between the start of year and end of year assessment, suggesting that children make progress on average in the first year of formal schooling. Of those present for both assessments, the progress scores for children were analysed to determine what proportion of assessed children make progress over the period. 1005 or 5.48% were omitted due to maximum scores on the first assessment and 1176 or 6.41% for minimum scores on the second occasion, to remove the impact of ceiling and floor effects. Of the children who participated in both the start and end of year assessments, only 1.60% failed to make any progress over the period between assessments.

Item Dependencies

One possible explanation of the findings given the nature of the methodology is that the order may result due to the nature of the selected numbers, their order of presentation and the use of a stopping rule after children answer incorrectly. As such, the order of numbers presented may be considered tautological. AUTHOR, (2018) used a range of methods (including mathematical simulations) to show that, in the English and Scottish data, this is not the case using the IPIPS assessment. They simulated scores for 300,001 students and difficulties for 21 items. These were used to calculate the probability of each individual getting each question correct. The stopping rules were then applied to simulate an assessment using the probabilities alone. Thus, new item difficulties were calculated. This process was repeated 1000 times and it was shown that there was a very close correspondence between the original and newly estimated item difficulties. Furthermore, the observed groupings of item difficulties remained constant. Although some numerals altered slightly in difficulty, these were within bands only; i.e. digits 3 and 4 may switch difficulties, but not 4 and 7, which crosses a band. It was concluded that there was no tautology.

In this study, data was examined to detect potential item dependencies. Andrich and Kreiner (2010) suggest that dependencies can be identified by examining the correlations between the residuals of observed and expected responses where larger correlations would suggest potential dependencies. Table 6 shows a tabulation of these correlations. The average correlation strength is around .001 (with a SD of .02). No moderate to large correlations were recorded. This analysis (showing results almost identical to AUTHOR, (2018)) suggests that localised item dependencies are not an issue in this data set.

Table 6: Average residual correlations between observed and expected responses by country

	Start of Year	**End of Year**

	Average standardised residual correlation for items	SD	Average standardised residual correlation for items	SD
Australia A	-0.002	0.046	-0.003	0.040
Australia B	-0.003	0.032	-0.002	0.029
England	-0.003	0.043	-0.002	0.024
N. Zealand	-0.004	0.028	-0.003	0.068
Scotland	-0.003	0.026	-0.002	0.022
Brazil	-0.053	0.154	-0.034	0.212
Serbia	-0.038	0.120	0.040	0.166
Slovenia	-0.034	0.096	0.028	0.080
W. Cape	-0.010	0.063	-0.009	0.080
Overall	-0.002	0.024	-0.001	0.026

Discussion

This study had four empirically testable hypotheses: unidimensionality, identifiable ability clusters, invariance, and progression during the first year. This section addresses each of these in turn before discussing number identification ability more broadly.

With regards to the first hypothesis, the results of the principal components analysis suggest that, across the countries studied, the assessment used in this study explain in excess of 70% of the available variance in the ability to identify numerals. Although this varies slightly from country to country, there is little evidence of substantive clustering in the residuals to suggest that some other, yet unidentified ability or trait is being measured by this assessment. The evidence therefore suggests that the ability to identify numerals could be a unidimensional construct, supporting the first hypothesis.

The second hypothesis suggested that there should be distinct and detectable ability bands. AUTHOR, (2018) suggested that these should occur at items 1-5, 6-9, teens, 2-digits and finally, 3-digits. The analysis of this international data set confirms that this appears to be the case, with these bands being clearly identifiable. Furthermore, this banding is clearly consistent across all examined countries, with very little evidence suggesting that the numbers within these bands cluster differently. Children internationally appear to learn to identify numerals in the same order. The second hypothesis appears well supported and generally agrees with existing literature on how children progress in number identification (AUTHOR, 2018; Mix et al., 2014; Wynn, 1992). Although there were some numbers that did appear to deviate from this structure (such as 100 and 300), these do not undermine the overall thesis of this paper. Some numbers, despite their complexity, appear easier to grasp than others. This likely reflects the greater regularity of these numbers in the developmental environment of children and their prominence in the wider environment as a whole rather than some great leap in understanding behalf of a child (Byrge et al., 2014; Mix et al., 2014). Alternatively, it may reflect lexicalisation of some common numbers, which are perhaps processed differently (Cohen et al., 1994).

The third hypothesis suggested that this potential unidimensional scale should be invariant across different groups. In this study, invariance by sex, language, country, additional language, and testing occasion was considered. Sex and additional language showed no evidence of meaningful differences on items. Occasion only showed DIF on one item, the number 4, well below the 5% suggested by the null hypothesis. Country and language showed some evidence of variance between some items in excess of the 5% that would be suggested by the null hypothesis. Items 73 and 281 appeared to show larger differences between countries. However, there is no apriori reason to suggest why these should differ and unless they were replicated, may represent chance here. If they were replicated, the difference may be due to age of children. In the countries that found these larger items easier, they were normally at least half a year older than their comparator on average. It may thus be that they had developed cognitively in a way that would make recognition of these symbols easier. However, these are generally small when one considers the range of the logit scale. It should also be noted that while some individual items do differ between countries and language, they do not differ to the extent that they fall into different ability bands.Numerals 4 and 5 may differ in difficulty but still fall within the same band, i.e. numeral 4 does not suddenly appear in the teens, or even the higher single digits. This is borne out clearly in Figure 3. Whilst item 12 showed evidence of DIF between the computer and paper versions of the assessment, this difference also was not large enough for it to appear in different bands between Group 1 and 2. As such, the general pathway in which children appear to learn numerals does not appear to differ substantively between countries, languages and methodologies. Taken together, the data presented are therefore generally supportive of the third hypothesis.

For the fourth and final hypothesis, and in line with AUTHOR, (2018), only a small number of children failed to make progress within the first year of formal education (<2%). Of this small number, some will have had odd results due to sampling errors, whilst some may genuinely have regressed, perhaps due to special educational needs and, or illness at the time of assessment. However, it is clear to see that children generally progress through the stages highlighted by this study and supports previous work (Chard et al., 2003; Litkowski et al., 2020) that suggests that children make much progress in this domain during the first school year.

Universality

Taken together, the evidence presented here is further evidence that the pathway to understanding numerals is likely a singular one through which all children navigate in the early years. This is an important finding although see the limitations section below. Although to some this may seem like a simple common-sense conclusion, why should number identification follow this universal pattern if all we are doing is mapping words to symbols? One may think that it represents the way children begin to count, each number increasing by a magnitude of one. However, were it that simple, then difficulty should ascend from 1 through 2, 3, 4, 5, 6, 7, 8 and so on. That is not the case in these results, with the difficulty order actually being 1, 3, 4, 2, 5, 7, 6, 0, 8 and so on; with other deviations from the standard count sequence throughout. This also would not really account for difficulty jumps between bands 1-5 and 6-9 given the linear increase in magnitude. Simplistic, "common sense" interpretations are therefore questionable.

This data cannot address why and how this is pathway would be universal. There are several possible explanations, however. An innate biological sequence through maturity of cognition is a tempting proposition but, one could easily point to potential learnability as an explanation, where a necessary sequence of learned facts and concepts must be consolidated before it becomes usable. Cross cultural similarities in how we teach mathematics is also a possible contender. More work would be required to explore this further which we discuss below in limitations and future work.

What we can focus more on here is the utility of this pathway. Given the suggested unidimensionality of the measure and its largely invariant nature, a measure of number identification (which on a Rasch scale, is interval in nature) may be an early universal yardstick on which all children internationally could be meaningfully compared. To recap, the scale does not appear to be unduly influenced by sex, languages, nations, additional languages and teaching practice in the first year of schooling and, as discussed earlier, number identification ability is also strongly predictive of later mathematical skills and difficulties (Gersten et al., 2005; Jordon et al., 2009; Lyons et al., 2014; Martin et al., 2014; AUTHOR, 1999). The implications for number identification as a practical classroom tool and a viable internationally comparative research tool cannot be overstated.

Learning to identify numerals

This research describes the general order of progression of numeral recognition. The difficulty jumps identified in Figures 1, 2 and 3 may provide some insight into how children are beginning to learn these numbers. Purpura et al. (2013) suggested that ‘numeral knowledge’ may be the bridge between informal and formal mathematics as more formal operations require connections to be made between early informal skills, such as counting, cardinality, quantity, and the written symbols used in more advanced mathematics. This may help explain why number identification is consistently one of the strongest predictors of future mathematics performance (Purpura et al., 2013; Lembke & Foegen, 2009). The difficulty jumps in number recognition may perhaps represent key points where informal mathematical knowledge is consolidated and assimilated into a wider ‘numeral knowledge’.

Doubtless the first digits that are recognised are simply symbols corresponding to words that the children have learned. Much like count sequences (Fuson, 1988; Gelman & Gallistel, 1978) there may or may not, at least in the very youngest of children, be any deeper understanding of numerosity. But as the numbers become more advanced, there surely is understanding developing with them as recent work by Litkowski et al., (2020) suggests. Their work attempted to illustrate trajectories of early informal skills such as one-to-one correspondence, cardinality and counting in particular, in relation to children’s ability to work with early numbers from 1 to 16 from ages 3 to 5.5 years. These informal skills are positively and moderately correlated with each other and with number identification ability. They also develop progressively with age as applied to higher numerals although, their study stops in the mid-teens. The ability to identify numerals perhaps represents the culmination of informal skills and the stages identified in this research may correspond to children’s progressive consolidation of them. The sample in this study. beginning at age 4 and progressing to age 7, suggests consolidation of some of these skills will still be taking place.

Small subsets of numbers such as arrays between 1 and 5 objects, can be processed very quickly from an early age via subitizing; a quantitative evaluation of small arrays without formal counting. This ability appears early in development and normally applies to arrays of around three objects (increasing to as many as five as they mature – Starkey & Cooper, 1995). Benoit et al., (2004) suggest that subitising is the mechanism by which children in fact acquire their first number words. As such, in Hindu-Arabic numerals, the first five symbols are the most widely used to recognise and describe small arrays in environments and are easily grasped by even very young children (see introduction). The understanding and competency in other informal mathematics abilities required to grasp these first symbols is likely to be comparatively low.

The difficulty gap transitioning from numerals 5 to 6, 7, 8, 9. is likely large because it potentially accompanies the transition from relying on subitising small arrays to the integration of other informal skills such as ono-to-one correspondence, verbal counting and cardinality. At this point, children need to begin making connections between different numerosity concepts to make sense of the count sequence relative to actual quantities, number words and symbol representations. This integration of skills increases the cognitive complexity of the task. The next large difficulty jump is to the early and later teens, where the Hindu-Arabic system has both a necessity to begin to understand some form of place-value but also, in some languages such as English, an irregular nomenclature e.g. where there is no one-teen, two-teen and three-teen, they become eleven, twelve and thirteen onwards. This likely requires additional learning. The jump between teens and two digits is less pronounced and likely reflects increasing consolidation of the early informal skills, as well as confidence in their use, and the beginnings of understanding rudimentary place value as they learn to decode more numbers. The jump between two and three digits is larger compared to teens and two digits, perhaps because the child needs to consolidate all these earlier principles effectively to scale them to three digits. Whilst not examined in this study, it would be a logical prediction that the jumps between 3 and 4 digits and beyond, would become successively smaller as children demonstrate mastery of the integration of these principles. As experience and exposure to numbers grow and children are interpreting them more regularly, confidence and competence in early numerical skills, such as decoding, (below) will likely make successive jumps to larger, unfamiliar numbers, easier.

Cognitive models of number decoding (Dotan & Friedmann, 2018; McCloskey et al., 1985) all centre upon syntactic and number-word frames in the process of identifying and producing a presented numeral. Single digits require a simple insertion from the lowest lexical class (‘Ones’). Decoding multi-digit numerals however requires understanding of the relation between classes to ultimately deduce the correct number word. For the teens it may also be that there is simple symbol recognition rather than actual decoding. ‘Lexicalising’, rather than decoding, some or potentially all of the numerals 11 through 19 is probably not impossible for young children and may even be a first strategy for children learning numerals prior to grasping early informal principles and place-value. This would allow some children to recognise some numerals without necessarily consolidating all the necessary informal skills to this point. But as two-digit numbers get higher, an understanding of place-value becomes essential for effective, accurate decoding. This continues into three digits and beyond. Whilst some larger numerals may be ‘lexicalised’ such as common benchmarks like 100 or environmental regularities such as a child’s house number, the vast majority will not be common across all children and decoding is the only way to interpret the majority of presented numerals.

Decoding processes, mastery of informal mathematical skills and increasingly sophisticated place-value understanding likely develop together in young children to make numeral identification possible (Litkowski et al., 2020). It should be noted however that the processes such as decoding have been simplified for discussion here and consists of multiple functionally independent sub and peripheral processes rather than singular centralised mechanisms (Dotan & Friedmann, 2018). A breakdown of any one or more of these could lead to identification or production errors, or cascade in such a way as to hamper the normal development of the identification process. As such, the pathway described in this paper likely applies to the normal development of Hindu-Arabic number identification and does not preclude the possible existence of additional or compensatory pathways emerging through developmental or acquired deficits to cognition.

Limitations

There are several limitations within this study that should be acknowledged. Firstly, the present analyses appear to suggest that local item independence is not problematic within this dataset, and this suggests the described pathway is not an artefact of the method. The use of stopping rules however does present limitations. It may mask some subtle elements of number identification. For instance, digits 6 and 9 are often harder to discriminate between (Gibson & Levin, 1975), and 0 is often understood later than other single digits (Wellman & Miller, 1986). Our analysis in fact bears this out, with these three numerals having higher difficulties than digits 1 to 5. As such, children may still struggle with these digits despite having knowledge of others. However, it is possible that children who identified them and one other numeral incorrectly, but could potentially identify other numbers, are exited from the assessment and are unable to demonstrate their knowledge. Differences in teens may also be masked. For example,11, 12 follow a different pattern to 13, 14. Future studies should consider implementing more liberal stopping rules to overcome this limitation. The most robust way of assessing the pathway without fear of tautology would be to give children all numbers 0 to 100 in a random order without any stopping rule. However, this is likely unfeasible both ethically and pragmatically, as forcing young children to continue when they are clearly making mistakes can lead to distress, and even without errors, boredom, and test fatigue. Unwieldy as they are, a stopping rule of some specification is a necessity.

Although we tentatively suggest that this pathway is universal, it must be acknowledged that the countries participating in this project have well developed, westernised education systems using Hindu-Arabic numerals assessed on entry to formalised education systems. Additional work is required however to examine if this pathway extends to additional languages and cultures, particularly with linguistic rule exceptions or different numerical structures. Testing this hypothesis on a large, nationally representative sample of individuals developing under different numeral systems would be elucidative, although this may be difficult given the almost universal adoption of Hindu-Arabic numerals across the globe. There are some cultural exceptions that may show that this pathway is limited. For instance, research has shown several different count systems in regions such as Papua New Guinea (Saxe, 2012). Would we expect our yardstick to work equally well for those developing in such systems? The explicit base-10 structure used in Chinese numerical language may also mean that the teens and general two-digit differences may not exist in this language. It must be acknowledged that different educational systems and practices (particularly systems that are not universally accessible or are less developed or formalised) may foster different patterns of learning in this domain. Although there is no specific apriori theoretical reason to suggest why this pathway, or the skills underpinning it, should vary, any true test of universality would need to consider this possibility. Similarly, differences in home circumstances or pre-school variability were also not accounted for due to differences in reporting methods across projects. Home background in the early years has a clear effect on early numeracy development (Jordan et al., 2007; 2009) and examining this pathway relative to socioeconomic factors in the early years would be informative. Interestingly however, socioeconomic background was examined in AUTHOR, (2018) and invariance of the pathway remained although this test is limited to only two countries. Pre-school variability has also been noted to be a factor in early informal mathematical skills development (Engel et al, 2016). Further exploration here is required as this pathway may not be replicable when comparing with nations that are very different in terms of educational culture and practice. Should the pathway be demonstrated however, this scale may provide further novel insights into how children are mentally using, representing, and mapping number.

Finally, data from these projects were not longitudinal beyond the first year in formal education. This limits the ability to track further progress with increased age and schooling, such as recognition of numbers of four or more digits, as well as prevents examining the relationship between the progress made and future attainment. However, as reviewed previously, there is a strong literature that suggests the predictive long-term effects of early number identification. This could also be extended into pre-school/kindergarten given that many children, as noted earlier, can recognise multi-digit numbers before entry into formal education. Are the ability jumps the same in younger children? Our model suggests they should not be, but very young children tend to subitise arrays of 1 to 3, with 4 and 5 being possible in older children (Starkey & Cooper, 1995). As such, it may be that in the very young, we see a further transition as we move from 3, to 4 and 5.

Future work

There are multiple research questions that could be considered from the findings presented within this study. Firstly, and perhaps most obviously, extending these findings to other language groups should be a priority although, without good reason to the contrary, we would anticipate confirmation of this pathway. Secondly, can the scale be used to compare the developmental levels of children within and across countries? If so, what does that developmental level relate to, for example, mathematical or cognitive development more generally, teaching and training, the language used for numbers? Thirdly, to what extent is the scale predictive of future success in literacy and numeracy? We reviewed earlier that number identification does appear predictive of these in the early years but only at specific time points (Chard et al., 2005; Jordan et al., 2009; AUTHOR, 1999). How predictive is this ability of later qualifications in secondary or tertiary education, requiring more longitudinal designs to answer? Fourthly, in what way does this number identification scale relate to the development of other facets of mathematical development, both concurrently and predictively? Litkowski et al., (2020) seems to suggest that informal skills are developing together but more testing and specific designs would be needed to answer this question more clearly. Finally, does the scale extend to older age groups, possibly even adults, and can it be used productively in studies on such populations?

Psychological and Educational Implications

While this finding of a pathway may initially be considered educationally obvious at face value to practitioners in the early years, understanding this pathway may be more beneficial on closer consideration. A simple assessment of numeral knowledge, taking minutes to administer, on entry to school can offer an insight into initial mathematical competency. Indeed, the IPIPS assessment from which this measure originates was intended to do just that alongside other competencies and provides useful information for teachers, illustrating ‘next steps’ in addition to simply identifying a child’s position in their learning.

Understanding that number identification occurs in stages, however, is also useful and knowing what numbers children can and cannot identify will give an insight into their learning and consolidation of related skills such as facets of informal mathematics and place value. Number based activities can thus be appropriately tailored to children based on which stage of number identification they are at, as this study suggests that children need to consolidate understanding of each step in turn whilst moving through the stages seems to get easier the more they learn. Many practitioners may be tempted to think that the ability jump between 25 and 35 is larger than the ability jump between 5 and 6 when this in fact is not so according to this data. Possessing such knowledge may encourage alternative ways of teaching number in the first year of school and may allow for greater classroom differentiation based upon on entry ability. Finally, research also shows that failure to grasp early numerical competencies such as number identification is an early sign of numerical processing difficulties in later education (Gersten et al., 2005). This is presumably because it represents a failure to consolidate those earlier informal mathematical skills (Purpura et al., 2013). As such, a simple number identification assessment at points in the first year may be a useful screener for practitioners. Children who are making little or no progress in number identification in the first year of school can be quickly identified for further specialist investigation and intervention. Our analysis shows that whilst the overwhelming number of pupils make progress in that first year, 1.6% do not. These children may be amongst the most at risk and require some form of intervention.

This study also provides tentative evidence for the validity of using the scale within national or international comparative frameworks. International assessments such as PISA, PIRLS, TIMSS are highly prized by researchers and policy makers alike despite the limitation that they do not have a baseline assessment to which progress can be mapped from (AUTHOR, 2004). A simple number identification paradigm could be used internationally to achieve such an end. In a similar vein, many national progress measures are similarly flawed (AUTHOR, 2004) because they lack information regarding a child’s ability on entry to formal education. As number identification ability appears invariant across many different demographic factors in this data at least, early assessment of this ability may assist schools and authorities in monitoring and tracking progress, and which may subsequently improve reporting and accountability metrics (which vary considerably internationally – AUTHOR, (2017)).

Conclusion

This investigation extends a more restricted paper based on two similar English-speaking countries using the same assessment (AUTHOR, 2018), and together these papers provide support for the view that there may be a universal developmental pathway, which children follow as they learn to identify number symbols. Number identification measures may thus prove a useful tool for both researchers and practitioners alike, particularly within an international context.

Disclosure Statement

No potential conflict of interest was reported by the authors.

Funding

This work did not receive funding support.

References

Baroody, A. J., Gannon, K. E., Berent, E., & Ginsburg, H. P. (1983). The development of basic formal math abilities. *Acta Paedologica, 1*(2), 133–150.

Baroody, A. J., & Wilkins, J. L. M. (1999). The development of informal counting, number and arithmetic skills and concepts. In Copley, J. V. (Ed). *Mathematics in the Early Years*, 48– 65. Washington, DC, National Association for the Education of Young Children.

Benford, F. (1938). The law of anomalous numbers. *Proceedings of the American Philosophical Society*, *78*(4), 551–572.

Benoit, L., Lehalle, H., & Jouen, F. (2004). Do young children acquire number words through subitizing or counting? *Cognitive Development*, *19*(3), 291-307. https://doi.org/10.1016/j.cogdev.2004.03.005

Benson, D. F., & Denckla, M. B. (1969). Verbal paraphasia as a source of calculation disturbance. *Archives of Neurology*, *21*(1), 96-102. https://doi.org/10.1001/archneur.1969.00480130110011

Bond, T., & Fox, C. M. (2015). *Applying the Rasch model: Fundamental measurement in the human sciences*. Routledge.

Burg, S. (2008). An investigation of dimensionality across grade levels and effects on vertical linking for elementary grade mathematics achievement tests. *Paper presented at National Council on Measurement in Education* (NCME) Annual Meeting, New York City, April 2008.

Byrge, L., Smith, L. B., & Mix, K. S. (2014). Beginnings of place value: How pre-schoolers write three-digit numbers. *Child Development*, *85*(2), 437–443. https://doi.org/10.1111/cdev.12162

Carey, S. (2004). Bootstrapping and the origins of concepts. *Daedalus*, *133*(1), 59–68. https://doi.org/10.1162/001152604772746701

Chard, D. C., Clarke, B., Baker, S., Otterstedt, J., Braun, D., & Katz, R. (2005). Using measures of number sense to screen for difficulties in mathematics: Preliminary findings. *Assessment for Effective Intervention*, *30*(2), 3–14. https://doi.org/10.1177/073724770503000202

ChartsBin statistics collector team (2009). Starting age of compulsory education around the World, *ChartsBin.com*, viewed 23rd May, 2018, http://chartsbin.com/view/z87.

Claessens, A., Engel, M., & Curran, E. C. (2014). Academic Content, Student Learning, and the Persistence of Preschool Effects. *American Educational Research Journal,* 51(2), 403-434. https://doi.org/10.3102/0002831213513634

Cohen, L., Dehaene, S., & Verstichel, P. (1994). Number words and non-number words: A case of deep dyslexia extending to Arabic numerals. *Brain*, *117*(2), 267-279. https://doi.org/10.1093/brain/117.2.267

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

DeLoache, J. S., de Mendoza, O. A. P., & Anderson, K. N. (1999). Multiple factors in early symbol use: Instructions, similarity, and age in understanding a symbol-referent relation. *Cognitive Development*, *14*(2), 299-312. https://doi.org/10.1016/s0885-2014(99)00006-4

Dotan, D., & Friedmann, N. (2018). A cognitive model for multidigit reading: Inferences from individuals with selective impairments. *Cortex, 101*, 249-281. https://doi.org/10.1016/j.cortex.2017.10.025

Engel, D. M., Claessens, A., & Finch, M. A. (2016). Mathematics content coverage and student learning in kindergarten. *Educational Researcher*, *45*(5), 293-300. https://doi.org/10.3102/0013189x16656841

Fuson, K. C. (1988). *Children's Counting and Concepts of Number*. New York, NY: Springer Verlag

Fuson, K. C. (1990). Conceptual structures for multiunit numbers: Implications for learning and teaching multidigit additions, subtraction and place value. *Cognition and Instruction*, *7*(4), 343–403. https://doi.org/10.1207/s1532690xci0704_4

Gelman, R., & Gallistel, R. C. (1978). *The Child's Understanding of Number*. Cambridge, MA: Harvard University Press.

Gersten, R., Jordan, N. C., & Flojo, J. R. (2005). Early identification and interventions for students with mathematics difficulties. *Journal of Learning Disabilities*, *38*(4), 293-304. https://doi.org/10.1177/00222194050380040301

Gibson, E. J., Levin, H. (1975). *The Psychology of Reading*. The MIT Press.

Jordan, N. C., Kaplan, D., Locuniak, M. N., & Ramineni, C. (2007). Predicting first-grade math achievement from developmental number sense trajectories. *Learning Disabilities Research & Practice*, *22*(3), 36–46. https://doi.org/10.1037/a0014939

Jordan, N. C., Kaplan, D. Ramineni, C., & Locuniak, M. N. (2009). Early math matters: Kindergarten number competence and later mathematics outcomes. *Developmental Psychology*, *45*, 850–867. doi:10.1037/a0014939.

Joseph, G. G. (2009). *A passage to infinity: Medieval Indian mathematics from Kerala and its impact*. SAGE Publications India.

Krajewski, K., & Schneider, W. (2009). Exploring the impact of phonological awareness, visual-spatial working memory, and preschool quantity-number competencies on mathematics achievement in elementary school: Findings from a 3-year, longitudinal study. *Journal of Experimental Child Psychology, 103*(4), 516–531. https://doi.org/10.1016/j.jecp.2009.03.009

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*(2), 395-438. https://doi.org/10.1016/j.cognition.2006.10.005

LeFevre, J. A., Fast, L., Smith-Chant, B. L., Skwarchuk, S. L., Bisanz, J., Kamawar, D., & Penner-Wilger, M. (2010). Pathways to mathematics: Longitudinal predictors of performance. *Child Development*, *81*(6), 1753-1767. https://doi.org/10.1111/j.1467-8624.2010.01508.x

Lembke, E., & Foegen, A. (2009). Identifying early numeracy indicators for kindergarten and first-grade students. *Learning Disabilities Research & Practice*, *24*(1), 12–20. https://doi.org/10.1111/j.1540-5826.2008.01273.x

Linacre, M. (2015). *Winsteps Rasch Measurement 3.90*. www.winsteps.com

Litkowski, E. C., Duncan, R. J., Logan, J. A. R., & Purpura, D. J., (2020). When do pre-schoolers learn specific mathematics skills? Mapping the development of early numeracy knowledge. *Journal of Experimental Child Psychology*, 195. Article 104846. https://doi.org/10.1016/j.jecp.2020.104846

Lyons, I. M., Price, G. R., Vaessen, A., Blomert, L., & Ansari, D. (2014). Numerical predictors of arithmetic success in grades 1–6. *Developmental Science*, *17*(5), 714-726. https://doi.org/10.1111/desc.12152

Martin, R. B., Cirino, P. T., Sharp, C., & Barnes, M. (2014). Number and counting skills in kindergarten as predictors of grade 1 mathematical skills. *Learning and Individual Differences*, *34*, 12–23. https://doi.org/10.1016/j.lindif.2014.05.006

Merkley, R., & Ansari, D. (2016). Why numerical symbols count in the development of mathematical skills: Evidence from brain and behaviour. *Current Opinion in Behavioral Science*, *10*, 14–20. https://doi.org/10.1016/j.cobeha.2016.04.006

Malofeeva, E., Day, J., Saco, X., Young, L., & Ciancio, D. (2004). Construction and evaluation of a number sense test with Head Start children. *Journal of Educational Psychology*, *96*(4), 648–659. https://doi.org/10.1037/0022-0663.96.4.648

McCloskey, M., Caramazza, A., & Basili, A. (1985). Cognitive mechanisms in number processing and calculations: Evidence from dyscalculia. *Brain and Cognition*, *4*(2), 171-196. https://doi.org/10.1016/0278-2626(85)90069-7

Miller, K. F., Kelly, M., & Zhou, X. (2005). Learning mathematics in China and the United States: Cross-cultural insights into the nature and the course of preschool mathematical development. In Campbell, J. I. D. (Ed), *Handbook of Mathematical Cognition*, pp 163-178. New York: Psychology Press.

Miller, K. F., Smith, C. M., Zhu, J., & Zhang, H. (1995). Preschool origins of cross-national differences in mathematical competence. *Psychological Science*, *6*(1), 56-60. https://doi.org/10.1111/j.1467-9280.1995.tb00305.x

Mix, K. S. (2009). How Spencer made number: First uses of the number words. *Journal of Experimental Child Psychology*, *102*(4), 427–444. https://doi.org/10.1016/j.jecp.2008.11.003

Mix, K. S., Huttenlocher, J., & Levine, S. C. (2002). *Quantitative development in infancy and early childhood*. New York: Oxford University Press

Mix, K. S., Prather, R. W., Smith, L. B., & Stockton, J. D. (2014). Young children's interpretation of multidigit number names: From emerging competence to mastery. *Child Development*, *85*(3), 1306–1319. https://doi.org/10.1111/cdev.12197

Mullis, I. V. S., Martin, M. O., Foy, P., & Hooper, M. (2016). *TIMSS 2015 International Results in Mathematics*. Retrieved from Boston College, TIMSS & PIRLS

International Study Center website: http://timssandpirls.bc.edu/timss2015/international-results/

National Mathematics Advisory Panel. (2008). *Foundations for success: Final report of the National Mathematics Advisory Panel*. Washington, DC, U.S. Department of Education.

Oberhuemer, P. (2005). International perspectives on early childhood curricula. *International Journal of Early Childhood*, *37*(1), 27-37. https://doi.org/10.1007/bf03165830

Pixner, S., Zuber, J., Hermanova, V., Kaufmann, L., Nuerk, H. C., & Moeller, K. (2011). One language, two number systems and many problems: Numerical cognition in the Czech language. *Research in Developmental Disabilities*, *32*(6), 2683-2689. https://doi.org/10.1016/j.ridd.2011.06.004

Purpura, D. J., Baroody, A. J., & Lonigan, C. J. (2013). The transition from informal to formal mathematical knowledge: Mediation by numeral knowledge. *Journal of Educational Psychology*, *105*(2), 453–464. https://doi.org/10.1037/a0031753

Purpura, D. J., & Ganley, C. M. (2014). Working memory and language: Skill-specific or domain-general relations to mathematics? *Journal of Experimental Child Psychology, 122*, 104–121. https://doi.org/10.1016/j.jecp.2013.12.009

Purpura, D. J., & Reid, E. E. (2016). Mathematics and language: Individual and group differences in mathematical language skills in young children. *Early Childhood Research Quarterly*, *36*, 259–268. https://doi.org/10.1016/j.ecresq.2015.12.020

Rasch, G. (1960). *Studies in mathematical psychology: I. Probabilistic models for some intelligence and attainment tests*. Danmarks Paedagogiske Institut.

Ryoo, J.H., Molfese, V.J., Brown, E.T., Karp, K.S., Welch, G.W., & Bovaird, J.A., (2015). Examining factor structures on the Test of Early Mathematics Ability — 3: A longitudinal approach. *Learning and Individual Differences*, 41, 21-29. https://doi.org/10.1016/j.lindif.2015.06.003

Sarama, J., & Clements, D. H. (2009). *Early Childhood Mathematics Education Research: Learning Trajectories for Young Children*. New York, NY: Routledge.

Saxe, G.B., (2012). *Cultural Development of Mathematical Ideas: Papua New Guinea Studies*. Cambridge University Press. https://doi.org/10.1017/CBO9781139045360

Seron, X., & Fayol, M. (1994). Number transcoding in children: A functional analysis. *British Journal of Developmental Psychology, 12*(3), 281-300. https://doi.org/10.1111/j.2044-835x.1994.tb00635.x

Siegler, R. S., & Booth, J. L. (2004). Development of numerical estimation in young children. *Child Development*, *75*(2), 428–444. https://doi.org/10.1111/j.1467-8624.2004.00684.x

Smith R. M. (2000). Fit Analysis in Latent Trait Measurement Models. *Journal of Applied Measurement*, *1*(2), 199-218.

Starkey, P., & Cooper, R. G., (1995). The development of subitizing in young children. *British Journal of Developmental Psychology*, *13*(4), 399-420.

Stenner, J. A., Simpson, M. A., Fisher, Jnr, W. P., & Burdick, D. S. (2015). A unified theory of task difficulty in K-12 mathematics. *Paper presented at the National Council on Measurement in Education Annual Meeting, Chicago,* April 2015.

TIMSS Encyclopaedia (2015). Retrieved from *Boston College, TIMSS & PIRLS International Study Center* website: http://timssandpirls.bc.edu/timss2015/international-results/encyclopedia/

Wagnar, S. H., & Walters, J. (1982). A longitudinal analysis of early number concepts. In *Action and Thought: From Sensorimotor Schemes to Symbolic Operations*, edited by Foreman, G. 137–161. New York: Academic.

Wellman, H. M. & Miller, O. F. (1986). Thinking about nothing: Development of concepts of zero. *British Journal of Developmental Psychology*, *4*(1*)* , 31-42. https://doi.org/10.1111/j.2044-835X.1986.tb00995.x

Wynn, K. (1992). Children's acquisition of the number words and the Counting System. *Cognitive Psychology*, *24*(2), 220–251. https://doi.org/10.1016/0010-0285(92)90008-P