Preschoolers' grasp of Cardinality:

A Methodological Examination and Theoretical Investigation

using Meta-Analysis, Experimentation

and Micro-Genetic Design.

Hannah Thomas, BSc., MSc.

This thesis is submitted

in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

June 2025

Department of Psychology

Declaration

I hereby declare that the work presented in this thesis has not been submitted, in whole or in part, for the award of a higher degree at this or any other university. I further declare that this thesis is a product of my own work, and the intellectual content of this thesis reflects my won thinking. All experimental studies included in this thesis were completed under the supervision of Professor Charlie Lewis and Dr Kirsty Dunn.

Studies one and three within this thesis are in preparation for publication:

Thomas, H., Lewis, C., & Dunn, K. (in preparation). A Systematic Review and Meta-analysis of the Give-N task as a measure of Cardinal Proficiency.

Thomas, H., Lewis, C., & Dunn, K. (in preparation). A Micro-Genetic Exploration of Cardinal Proficiency Development: Sudden vs Gradual Change.

Hannah Thomas, BSc., MSc.

Lancaster University, UK

Statement of Authorship

Experimental Chapter 1: A Systematic Review and Meta-analysis of the Give-N task as a measure of cardinal proficiency
Conception and design of study: Hannah Thomas, Charlie Lewis, Kirsty Dunn
Acquisition of data: Hannah Thomas
Data analysis: Hannah Thomas
Data interpretation: Hannah Thomas, Charlie Lewis, Kirsty Dunn
Drafting the manuscript: Hannah Thomas
Revising the manuscript: Hannah Thomas, Charlie Lewis, Kirsty Dunn
Contribution of principal author: 85%

Experimental Chapter 2: Examining the Influence of Pretend Play on Children's Displays of Cardinal Proficiency
Conception and design of study: Hannah Thomas, Charlie Lewis, Kirsty Dunn
Acquisition of data: Hannah Thomas
Data analysis: Hannah Thomas
Data interpretation: Hannah Thomas, Charlie Lewis, Kirsty Dunn
Drafting the manuscript: Hannah Thomas
Revising the manuscript: Hannah Thomas, Charlie Lewis, Kirsty Dunn
Contribution of principal author: 85%

Experimental Chapter 3: A Micro-Genetic Exploration of Cardinal Proficiency Development: Sudden vs Gradual Change
Conception and design of study: Hannah Thomas, Charlie Lewis, Kirsty Dunn
Acquisition of data: Hannah Thomas
Data analysis: Hannah Thomas, Charlie Lewis
Data interpretation: Hannah Thomas, Charlie Lewis, Kirsty Dunn
Drafting the manuscript: Hannah Thomas
Revising the manuscript: Hannah Thomas, Charlie Lewis, Kirsty Dunn
Contribution of principal author: 85%

We the undersigned agree with the above stated "proportion of work undertaken" for each of the above chapters contributing to this thesis.

Hannah Thomas:

Charlie Lewis:

Kirsty Dunn:

Acknowledgements

I would like to thank my supervisors, Professor Charlie Lewis and Dr Kirsty Dunn, for all their help and support over the last five years. Without them this thesis would not have been possible. The combination of theoretical and practical advice has been instrumental to this thesis, and for that I am forever grateful.

I would also like to thank Lancaster University Babylab, who again, have supported my journey both practically and emotionally. I am, again, eternally grateful. Likewise, thank you to all the amazing children and nursery settings who have made me feel incredibly welcome and have always greeted me with a smile.

Finally, I would like to thank my family, and my colleagues / friends from Kendal College. Their patience has been unwavering, and their support has no doubt ‘got me through’.

List of Figures and Tables

List of Figures

Experimental Chapter 1

Experimental Chapter 2

Experimental Chapter 3

List of Tables

Experimental Chapter 1

Experimental Chapter 2

Experimental Chapter 3

Table of contents

Abstract

Acquiring the cardinal principle is a foundational step in young children's mathematical development. It is at this point that children understand that the final number in a count represents the total number of items in that set. From this point children are able to manipulate numbers; add up, take away and multiply. Development of the cardinal principle is therefore the gateway to a conceptual understanding of number. Importantly, improving children's understanding of this principle contributes to higher mathematical competence, and this advantage is maintained throughout their education. It is therefore crucial that we understand how this skill develops. Such understanding enhances theoretical accounts, improves empirical studies and enables caregivers and practitioners to provide appropriate scaffolding to facilitate development. This thesis examines the primary measure of cardinal proficiency development, the Give-N task, and explores the pattern of development as this key skill is acquired. The first empirical chapter is a systematic review and meta-analysis. It suggests that the Give-N task may not be a reliable measure of cardinal proficiency and highlights the multitude of ways this task is implemented throughout the developmental literature. Experimental chapter two expands upon this by manipulating one key aspect of the task. Here, we find that small changes in the way the Give-N task is implemented may affect children's displays of cardinal proficiency. The final experimental chapter reports, in a micro-genetic study over eight consecutive weeks of testing, the complexity of how cardinal proficiency develops. The main finding is that children's pattern of development is highly variable as they wrangle with conceptual understanding. Overall, this thesis highlights important methodological considerations when measuring cardinal proficiency and contributes to current theoretical understanding of how this pivotal skill develops. Our conclusions correspond with emerging evidence and suggest that development of the cardinal principle is more variable than the dominant theoretical accounts suggest. Children go

through a long period of change in which they wrestle with ideas as they gradually acquire knowledge, and conceptual insight is different in each individual.

Introduction Chapter

Cardinal proficiency is the gateway for children's understanding of number. When children understand this principle, they recognise that the final number in the count represents the total number of items in a set (Wynn 1990,1992). Knowledge of this principle is the foundation upon which children can then add, take-away, divide, and multiply numbers. This makes the cardinal principle a primary, integral, part of children's understanding of number, arithmetic, and mathematical concepts. Consequently, improving children's understanding of this principle contributes to higher mathematical competence, and this advantage is maintained throughout their education (Aunio & Niemivirta, 2010; Chu et al., 2015; Chu et al., 2016; Jordon et al., 2009; Koponen et al., 2013).

Given that recognising that the cardinal principle is a crucial milestone for children, it is imperative that we understand how this skill emerges. This can enable practitioners and caregivers to provide appropriate scaffolding to facilitate development, and nurture young children's confidence when learning how to use number. To understand how this skill develops we must have testable theories, reliable measures, and standardised methods. Initial scoping of the literature suggests that conventional methods are not standardised or reliable and, whilst we may have testable theories, these have been static for over twenty years. Research has actively attempted to test these theoretical positions. However, the focus becomes a case of supporting or disproving specific aspects of theory, rather than exploring the foundational concepts upon which theory was constructed. As such, theoretical stances have become entrenched. Correspondingly, methodologies have become fixed, with the Give-N task deemed by many to be the acid test of cardinal proficiency, being unchanged since 1990. Of concern is the lack of studies exploring the consistency and reliability of this test.

The objective of this research is to explore the reliability and consistency of the Give-N task and to reflect upon whether it supports current theory. By examining the way in which the Give-N procedure works, we will consider the theoretical thinking around cardinal proficiency development, aiming to view theory through the lens of *change* and explore cardinal principle development in detail. Close analysis of children's performance will, hopefully, initiate a shift in epistemological stance, from static states to that of variability and change. The first study presented is a systematic review and meta-analysis of 113 studies using the Give-N task. Here, the reliability of the task is assessed along with the consistency of how this task is implemented. The second study is informed by these findings, and experimentally measures the variability in children's displays of cardinal proficiency evoked by a different implementation of the Give-N task. The third study focuses on variability and change, and adopts a micro-genetic method to explore, in detail, how cardinal proficiency develops. Testing weekly over three months allows analysis of the strategies children use when gaining cardinal proficiency, therefore enabling exploration of current theoretical positions. The objective is to generate more robust measures, alternative methodologies, and enhance our current theoretical understanding of how cardinal proficiency develops.

Give-N Task

The Give-N task is deemed by many to be the acid test of cardinal proficiency development. The task requires children to give or produce a sub-set of items from a larger set. Children are deemed to have passed the task and are credited with cardinal proficiency when they are able to give five or six (or more) items. Importantly, the Give-N task is ubiquitous throughout the developmental literature, and features heavily in the discontinuity vs core knowledge debate in cardinal principle development (Gelman & Gallistel, 1978; Le Corre et al., 2006; Sarnecka & Lee, 2009; Sarnecka & Carey, 2008; Wynn 1990, 1992). Likewise, the Give-N task is prominent when associating the Approximate Number System to cardinal proficiency and later arithmetical development (Dehaene, 2011; Hyde et al., 2017; Wilkey & Ansari, 2020).

More specifically, the Give-N task has been influential when advancing our theoretical stance regarding cardinal proficiency development. The prominent knower-levels account (Sarnecka & Carey, 2008; Wynn 1990, 1992) of how cardinal proficiency develops utilises the Give-N task, and in many ways this task and the account are inextricably linked. Furthermore, the Give-N task is a primary measure in other fields of developmental psychology and has advanced our knowledge of how early number skills are associated with language development (Purpura & Ganley, 2014), problem solving (Chu et al., 2018) and spatial skills (Verdine et al., 2014).

With such reliance on an individual task it is vital the Give-N task is robust, and this drives our research questions in chapter one and two.

Early Number Skills

The development of symbolic number begins prior to formal education, in early childhood. Long before children are given any direct instruction with mathematics or symbolic numbers, they are gaining knowledge of the count sequence through nursery rhymes and daily activities, and they start to understand that counting and numbers are important (Ginsburg et al., 2006; Litkowski et al., 2020). It is widely accepted that early proficiency with symbolic number predicts later mathematical achievement (Aunio & Niemivirta, 2010; Chu et al., 2015; Chu et al., 2016; Jordon et al., 2009; Koponen et al., 2013; Nguyen et al., 2016), highlighting the importance of these early skills.

There are several well documented counting principles that children acquire before they have a complete understanding of symbolic number. These are the ability to recite the number sequence, ordinality, one to one correspondence, subitizing, and cardinality (Gelman & Gallistel, 1978; Litowski et al., 2020; Sarama & Clements, 2009, 2014). Traditionally, it has been proposed that these develop in order, as above. Whilst this is accurate, it is important to note that these principles may be developing simultaneously.

Reciting the Number Sequence

This principle usually develops around the age of 18 months to two years, and is the ability to verbalise the count list i.e., count 'one, two, three, four, five'. This begins with singing songs that include number words and progresses to children being able to verbally count up to 10 or beyond. Critically however, at this stage, children do not understand the meaning of counting or numbers. The count list they are reciting could easily be 'A, B, C, D' as opposed to 'one, two, three, four'. The ability to verbally count or recite the number sequence is simply being able to articulate a rote learned count list (Sarama & Clements, 2009).

Ordinality

The principle of ordinality is understood when children recognise that the numbers in the count list have fixed positions, and are in a sequence -, i.e., two always comes after one, and five after four (Coles, 2014). This principle provides the foundation for understanding that larger numbers are later in the count list and smaller numbers appear earlier. Here, children begin to understand that numbers may have meaning, beyond that of a rote list, and children begin to think more deeply about the relationships between numbers (Coles, 2014).

One to One Correspondence

Traditionally, one to one correspondence has been identified as the next stage in the early number skills trajectory, and develops around the age of three. Here, children understand that when counting a set of objects each number in the count list only refers to one object. They understand that counting one object twice (i.e., labelling one object with number words two and three) would be an error, and missing an object when counting would also be an error (Gelman & Gallistel, 1978). Again, one to one correspondence is an important principle as children appreciate that counting may have meaning and therefore this needs to be accurate.

Subitizing

This is the ability to recognise rapidly the number of items within a set, without counting (Litkowski, 2020). Typically, pattern recognition facilitates this skill, and young children are only able to recognise sets of up to four items. Again, this development of this skill suggests that children are beginning to understand that numbers have meaning and can represent sets of items.

Cardinality

Cardinality or development of the cardinal principle usually happens around the age of three to four years. This is when children recognise that the final number when counting represents the total number of items in a set (Wynn 1990,1992). Knowledge of this principle is the foundation from which children can then add, take-away, divide, and multiply numbers. This makes the cardinal principle a primary, integral, part of children's understanding of number, arithmetic, and mathematical concepts. Once children have grasped the cardinal principle it is suggested they have undergone a conceptual change and now have a full grasp of symbolic number (Sarnecka & Carey, 2008). Development of this early numerical skill is the gateway to understanding formal mathematics and is vital in children's mathematical learning trajectory (Sarama & Clements, 2009). It is consequently the focus of this research project.

Current Theoretical Views of Cardinal Proficiency Development: Revise or Reinvent?

Background

Early accounts of cardinal proficiency development suggest a principles-first emergence of this skill. These accounts assert that children have an innate, implicit knowledge of counting principles and numerical understanding. However, they are unable to demonstrate these skills due to due procedural demands of tasks (principle (Gallistel & Gelman, 1992; Gelman, 1993; Gelman & Meck, 1983). Whilst this view is still held by some, a larger number of researchers claim that young children have no knowledge of symbolic number. Young children have some numerical knowledge, but this is non-symbolic and refers to their ability to differentiate between different ratios of dot arrays or series of events (Krajsci & Reynvoet, 2023). Whilst there is still debate regarding the foundations of symbolic number (i.e., does children's non-symbolic number knowledge represent innate numerical principles and underpin development of symbolic knowledge, or are they two distinct systems?), there is advancing research that non-symbolic and symbolic number are two distinct systems. Wilkey and Ansari (2000) synthesised the evidence to suggest these systems follow their own developmental trajectories and each system has its own distinct neural mechanism. Given these findings, there is a lack of consistent association between non-symbolic and symbolic number, and recent acknowledgment of confounds within non-symbolic research (Gilmore et al., 2014; Szűcs et al., 2013), there is movement away from a principles first view of numerical abilities. Furthermore, large amounts of published research have explored the principles first vs principles after debate with no definitive answer. The field of numerical cognition has recognised the need to suspend these discussions, to allow progression and growth. With these two points in mind, a greater number of researchers now focus on the principles after and knower-levels view of cardinal proficiency development.

The knower-levels account asserts that children progress incrementally through levels of development and, using the Give-N task as a primary measure, their description of how children develop an understanding of symbolic number is as follows. When children can reliably give one item, but not two items, they are deemed to be ‘one-knowers’. After a few months children progress and are able to reliably give two items, but not three, therefore they are deemed to be ‘two-knowers’. Children progress through the knower levels incrementally until they are able to reliably give five items. At this point a child is thought to be a cardinal principle knower and can reliably give all the numbers within their count list (i.e., as high as they can count), and their strategy to pass the task changes from grabbing the items to counting them out. Due to children’s displays of accurate ‘giving’ on this task, the knower-levels account asserts that when children can give five items, they have undergone a conceptual change and now understand the purpose of counting. They recognise that the last number word in a count represents the quantity in the set and therefore they have an understanding of symbolic number. As such, one, two, three, and four-knowers are categorised as sub-set knowers and five-knowers are cardinal principle knowers.

As discussed earlier, it is evident that this account of cardinal principle development is inextricably bound by the current measures which are employed. The Give-N task forms the basis of the theoretical description, and the assumptions of the theoretical account shape the constraints of the task. For example, numbers higher than five are typically not requested, as it is assumed that once a child can give five, they are cardinal principle knowers. Furthermore, whilst some of this description is indisputable, some parts are not supported empirically (this will be discussed in more detail in the following section). Consequently, the inference that children undergo a conceptual change when they are able to give five items may not be accurate. The reliance and interplay between the Give-N task and this dominant account of cardinal proficiency development necessitates rigorous analysis of both the Give-N task and these theoretical assumptions. We must assess if the Give-N task is robust and reliable, and we must use different measures to triangulate the findings that underpin current theory.

Analysis

The central tenet of the knower-levels account of cardinal proficiency development is: Children treat number words as mutually exclusive and therefore learn number words one by one (incrementally) until they are able to give five items, at which point they are able to give all the numbers in their count list. When children are unable to give five (or more) items they are sub-set knowers and do not understand symbolic number, whereas children who are able to give five (or more) items are conceptually advanced and understand the purpose of counting and therefore have grasped the cardinal principle (Lee & Sarnecka, 2010; Sarnecka & Gelman, 2004; Sarnecka & Lee, 2009; Wynn, 1900).

The account offers three key pieces of evidence to substantiate this tenet. Firstly, as evidence for incremental learning, the account discusses that children do not give 'known' numbers for numbers they do not yet know (Sarnecka & Gelman, 2004). Secondly, as soon as children can give five items (and therefore grasp the cardinal principle) they generalise this knowledge and are able to give all numbers within their known count list (i.e., all the numbers they can recite), evidencing a conceptual change when children are able to give five items (Carey 2004; Wynn, 1990; Wynn 1992). Thirdly, subset knowers grab and give items, whereas cardinal principle knowers count out items, again, as evidence for conceptual change when children can give five items (Chetland & Fluck; Wynn 1992). These key pieces of evidence will be explored below.

Children do not give 'known' numbers. It is suggested that once children have learnt a particular number word (e.g., three) and are therefore categorised as a three-knower they would not give three items when asked for an unknown number, such as five. Proponents of the knower-levels account use this mutually exclusivity as evidence to support their suggestion that sub-set knowers learn numbers one by one, and in order (Sarnecka &

Lee, 2009). Once again, the interplay between the knower-levels theory and the methodology used to support their claims is apparent. The suggestions within the knower-levels account have shaped how the Give-N task is implemented. It has become standard practice that if a child is able, for example, to give three items correctly, but then gives three for any other number, they are not categorised as a three-knower due to their errors rather than their success (Sarnecka & Gelman, 2004; Wynn 1992). The scoring criteria are heavily influenced by theory. The problem is the theoretical assumption that children are adhering to the mutual exclusivity rule when giving numbers during this task. Children may provide known numbers for unknown numbers due to familiarity; when faced with a task they cannot solve they stick with what they know, even when this might be wrong. For example, a child categorised as a three-knower may persistently give three items when asked for four, five or six items. They know how many items are in a set of three, and therefore give three items for the unknown sets of four or five when asked. Alternatively, perhaps children cannot disengage from a correct response - i.e., children repeatedly give two items following their correct response when asked to give two. With this reasoning children may be two-knowers. However, with the mutual exclusivity scoring criterion discussed above children are not credited with this knowledge. With this criterion there may be significant underestimation of children's abilities. This could mean the Give-N task, implemented as per the knower-levels theory, may be an unreliable measure.

The important point here is the lack of evidence to support either interpretation of children's behaviour. This is why there is a need to explore children's development of the cardinal principle using different methods, not just the Give-N task. The knower-levels theory may be correct, but without further evidence this cannot be verified. Again, the reliability of the Give-N task and the effects of implementation need assessing. Likewise, the principal assumptions of the knower-levels theory need validating. The interplay between theory and measure must be disentangled.

Generalisability. This assertion is founded on the assumption that conceptual advancement happens when children are able to give five items. This is a step change in children's development and therefore they are able to generalise their knowledge to all other known numbers. However, in the seminal study by Wynn (1990) children are not asked for numbers higher than six, and in the follow up study by Wynn (1992) children are not asked for numbers higher than five. This has become standard practice with the Give-N task. Evidently, without asking for higher numbers how can the generalisability claim be substantiated? Recent evidence is emerging that suggests that six or even seven knowers have not generalised the cardinal principle to all the numbers in their count list (Gunderson et al., 2015; Krajcsi & Fintor, 2023; Mussolin et al., 2012; Posid & Cordes, 2018; Wagner & Johnson, 2011). The inherent problem throughout the literature is that claims are made regarding higher numbers, when in fact numbers sets over five or six were not requested. The assumption that the cardinal principle is automatically understood when children can give five items is pervasive, and this constrains the methods.

Many different studies in support of generalisability and a discontinuous conceptual step-change make claims about unknown numbers without asking for numbers higher than five or six (e.g., Sarnecka & Carey, 2008; Sarnecka & Gelman, 2004; Sarnecka & Lee, 2009). This omission limits our knowledge of how children perform with larger unknown numbers and, again, compromises whether these studies can claim generalisability when numbers above five have not been requested regardless of children's successes. In addition, the incremental method of asking for numbers of items further constrains our knowledge. For example, if children fail on four items, five items are not requested, but how do we know they would not succeed with five items? If this has not been tested empirically then we cannot draw any firm conclusions. Once again, the methods utilised to measure cardinal proficiency are constraining our understanding of how this important skill develops, and current theoretical accounts need to reflect upon this.

Grabbing vs Counting. Here, the analysis of strategies to support theory is very coarse. We only have two categories, which cannot be representative of children's complex behaviours during any form of numerical task. Studies exploring strategy use during other mathematical achievements document many different strategy categories (e.g., Crooks & Alibali, 2014; Sahin et al., 2020), and whilst these studies are typically conducted with older children, there is no evidence to suggest that younger children's strategies would be any less varied. In fact, it could be argued that younger children may have use a greater variety of strategies, as they are less influenced by formal education. Furthermore, studies that focus on change and conceptual development document high variability in strategy use during periods of change (Adolph et al., 2008). The coarse use of strategies to support the knower-levels account suggests inadequate, imprecise measurement, thus reducing the utility of this evidence as support for the knower-levels account. Furthermore, the assertion that a grabbing strategy represents a lack of cardinal principle knowledge, and a counting strategy represents knowledge of this principle needs to be explored. This seems to be a widely accepted perspective. However, only a small number of studies have examined this. Further research is required to allow us to consider the mechanisms underpinning strategy changes and reflect upon whether these two different strategies do represent different levels of conceptual knowledge.

In addition, we have the same problem as with generalisability. How do we know what strategies children may use for larger numbers, when these are not tested. If we were to abandon the incremental method of the Give-N task and ask a child to give seven items (even if they failed to give five items), how do we know they would not adopt a counting strategy?

Some of the claims made by the knower-levels account are un-evidenced and need to be re-examined and explored using alternative methods.

Revise or Reinvent?

It is clear from the above discussion that the theory has influenced the measure, and the measure has influenced the theory. This interplay necessitates assessment of the Give-N task as a robust measure of cardinal proficiency, the use of different measures to triangulate current research findings, and the confines of this measure upon current theoretical accounts need to be removed. This does not mean we need to reinvent a whole new theory of cardinal proficiency development, but we do need to disengage from current theoretical assumptions and be open to revise the theory in light of alternative evidence.

More specifically, we need to re-examine the generalisability claim, and the coarse strategy categorisation that has been used. We should explore what children can and cannot do with larger numbers (over five), and the strategies children use when giving larger sets of items. In addition, we need to assess if the assumed conceptual advancement of the cardinal principle does take place when children can give five items. Emerging evidence suggests that generalisability may not automatically occur when children can give five items, and that children may be able to give seven items but not understand the cardinal principle (Krajsci & Reynvoet, 2024). Within this pivotal study the authors assessed children's number knowledge using the Give-N task and a verbal comparison task (i.e., "Mark picked four apples, and Emma picked seven apples, who picked more?"). Two important findings are reported. Firstly, the Give-N task revealed some children were able to give numbers higher than five (e.g., seven or eight items) but not all the items in their count list, violating the generalisation tenet within the knower-levels theory. Secondly, these children had similar performance on the comparison task, as those who were unable to give five or more items.

Specifically, sub-set knowers (those not able to give five or more) and children who were able to give six, seven or eight (but not all the numbers within their count list) were only able to compare ‘known’ numbers i.e., numbers they were able to give during the Give-N task. The authors conclude that, due to their lack of generalisation and their comparative performance matching that of sub-set knowers, there may be an additional category of children. Those able to give six, seven or eight etc (but not able to generalise) may be larger number sub-set knowers, and these children are still conceptually sub-set knowers and not cardinal principle knowers (Krajsci & Reynvoet, 2023). This is valuable evidence to suggest that children may not automatically understand the cardinal principle when they are able to give five items, and this supports a more gradual, individual, developmental trajectory of the cardinal principle.

This line of reasoning is beginning to question the current confines of theoretical thinking around cardinal proficiency development, and it is an important study to detail within this thesis. Whilst the study was published after commencement of this research, it complements our original aims and provides support for our later findings.

How do we Measure Cardinal Proficiency and is this Task a Robust Measure?

The Give a Number or Give-N task is currently the most used measure of cardinal proficiency development. The use of this task is ubiquitous throughout the developmental literature and appears in many different areas of research, not just numerical cognition. For example, the test is used in studies of language and working memory abilities (Purpura & Ganley, 2014), sharing capabilities (Chernyak et al., 2016; Chernyak et al., 2019), selective attention (Brueggemann & Gable, 2018), problem solving (Chu et al., 2018), and spatial skills (Verdine et al., 2014). Likewise, the Give-N task has, almost solely, been employed when testing theoretical positions surrounding cardinal proficiency development (Gelman &

Gallistel, 1978; Le Corre et al., 2006; Sarnecka & Lee, 2009; Sarnecka & Carey, 2008; Wynn, 1990, 1992). This has happened despite the relative paucity of studies examining its reliability and validity. With this issue in mind, it is imperative to test if this procedure is robust.

The standard task requires children to select numbers of items (e.g., two / three / four) from a larger set of items (usually between 10 and 15). Typically, a titration method is used whereby children are asked to give one item, then two items and so on, until they fail to give the correct number. At this point the child would be asked for the number below. If they succeed with the lower number they will be given another chance to provide the correct amount for the number above. If they fail to give this number on their second chance, the task ends and children are given a knower level equivalent to the highest number they are able to give (Wynn, 1990). For example, after giving two, three, and four items correctly, a child would then be asked to give five items. If they fail to produce five items, they would be asked to give four items again, if four items were given, a request for five items would be presented (second chance). If they fail again to provide five items, they would be deemed a 'four-knower'. Importantly, when a child is reliably able to give five items they are deemed cardinal principle knowers. They now are considered to understand the concept of symbolic number and the purpose of counting, and therefore no higher numbers are requested. This assumption is based upon the idea that sets up to a total of four are subitizable, while those of five and above are not (Lee & Sarnecka, 2009; Wynn, 1990).

On the surface this appears to be a straightforward, easy to implement task that measures what it should. Likewise, it corresponds with the dominant knower-levels theory of cardinal proficiency development (Lee & Sarnecka, 2010; Sarnecka & Gelman, 2004; Sarnecka & Lee, 2009; Wynn, 1990). Hence, it is understandable why this method has been widely adopted. However, whilst this methodology appears to have been standardised, its robustness has not been established empirically. This is witnessed in the different ways in which this task has been implemented, and this variation may affect children's displays of number knowledge. This has not yet been investigated. Furthermore, unless convergent validity between the variations has been established (Campbell & Fiske, 1959), only utilising one measure, in a very specific way, may be constraining our understanding of the exact concept we are trying to measure. For example, by not asking for numbers higher than five, or not asking for higher numbers once a child has failed twice, we may be limiting our knowledge of what children may or may not understand about larger numbers (Krajcsi & Fintor, 2023; Krajcsi & Reynvoet, 2023). As a result, theoretical insight may be compromised by the assumption that once children can give five, they have gained an understanding of the cardinal principle, and the task we use to measure this does not allow exploration outside of this assumption.

It is necessary to examine whether there are different ways in which this task has been implemented, if this variation affects children's displays of cardinal proficiency, and why a particular implementation may affect children's ability to succeed. This would throw light on whether the task needs to be standardised, if it is robust, and if current theoretical positions can account for any variations. Furthermore, the sole use of this 'go-to' task maybe limiting theoretical explanations of development. Such a possibility needs further consideration. Again, it seems timely to examine if other, less traditional, measures of cardinal proficiency, such as strategic variability, validate current theory.

Different Ways to Measure Cardinal Proficiency

There are several other tasks that can be used measure cardinal proficiency. Many of these closely correspond with the Give-N task, and only diverge as the procedural demands vary from task to task. Additionally, these measures have their own inherent methodological problems, and are used much less frequently within the literature. All the measures, including the Give-N task, assess cardinal proficiency numerically, and prescribe that cardinal proficiency is assessed on a 'pass' or 'fail' basis, when children can count, give, or recognise five items. Theory becomes bounded by this premise. The need to explore different ways of measuring cardinal proficiency development is again explicit. Do different measures affect children's displays of cardinal proficiency? Can tasks that do not focus on pass vs fail or numerical responses reveal a more gradual trajectory of development (Krajcsi & Fintor, 2023; Krajcsi & Reynvoet, 2023)

How Many Task

During this task children are shown picture cards with different numbers of animals or items on (e.g., three pigs). Firstly, they are asked, 'can you count the pigs?' and then 'so, how many pigs were there?'. Typically, after children have counted, the card is turned over before children are asked the how many question. This removes the option for children to just re-count the pigs, as the how many question can infer a re-count is required, or children may just re-count because the pigs are in sight (Muldoon et al., 2005).

Arguably, this task is less cognitively demanding than the Give-N task. Children do not have to keep a number in mind whilst they count out a set, thus reducing the working memory load. Likewise, children are instructed to count; they are given the correct strategy for this task, thus facilitating success. At the same time, this task has been criticised as children may have learnt the last word response rule. Rather than fully understanding the cardinal principle, they may just repeat the last number in the count as a rote response (Baroody et al., 2017).

In sum, the how many task is less procedurally demanding and may over-estimate children's cardinal principle knowledge. However, and more importantly, this task has the same potential inherent flaws as the Give-N task. The task ends when children fail, and the premise of the task is numerical. It only assesses the numbers given or spoken by the child. This may seem an obvious measure as we are, after all, assessing children's numerical competency. However, this assumption is likely to be both simplistic and reductionist. There are many other ways to assess children's understanding and it is imperative that other measures, including those from other fields of developmental psychology, are explored. Current theory cannot be validated or enhanced, when the tasks used to assess cardinal proficiency all use a similar and possibly simplistic measure of development.

Point to X

During this task children are shown picture cards with varying numbers of items or animals on each half of the page (Wynn, 1992). For example, three pigs on one half of the page and five pigs on the other half. The page is usually separated with a black line down the middle. Children are then asked to point to the side with 'x' number of animals / items. This task does not provide the children with a counting strategy, thus potentially improving the validity of this measure. However, it may be that children are comparing size, or surface area, or relying on in-exact numbers to solve the question successfully as in ANS studies (e.g., Clayton, Gilmore & Inglis, 2015). If they know five is later in the count list and therefore a bigger number, they could pick the corresponding 'larger' side without counting and fully understanding the cardinal principle. Furthermore, as with the how many task, assessment is pass vs fail and numerical, possibly constraining our understanding of cardinal principle development.

Comparison / Equivalence Tasks

The judgements around these tasks are mixed. Some argue that sharing and equivalence follow cardinal principle development (Baroody et al., 2017; Litowski et al., 2020). Whereas others see set comparison as a window into children's understanding of the cardinal principle and until numbers can be compared, children's conceptual understanding of numbers is limited (Krajcsi & Reynvoet, 2023; Muldoon et al., 2009; Piaget 1952). Tasks in which children make comparisons between set sizes, usually require children to make a judgement, 'Adam has three apples, Jane has five apples, who has more?', with visual aids or objects typically used alongside. Although these tasks are still numerically bound, they do remove the pass / fail constraint and begin to explore the concept of cardinal proficiency from a different stance. Rather than cardinality being inextricably linked to how many items a child can count or give, this creates an opportunity to examine the concept of cardinal proficiency through comparison. If a child understands that the final number in a count represents the total number of items in a set, then two different counts must identify different sized sets. Viewing cardinal principle development through this lens allows further exploration of current theory. Are 'traditional' non-cardinal principle knowers (those unable to give five or more items) unable to make *any* comparison judgements? Or are they able to compare known numbers? If they are able to make such comparisons, this may suggest that development of the cardinal principle is a gradual change and progression of knowledge. Children's understanding of the cardinal principle may begin to emerge before they are able to give five items, with this knowledge gradually becoming more secure and complete, with generalisable knowledge of the cardinal principle developing at different points for different children. As discussed earlier, evidence to support a gradual view of development is emerging, and these recent pivotal studies adopt comparison judgements to explore this. (Krajcsi & Reynvoet, 2023). This demonstrates the need for alternative methodologies and for the field to move away from numerical, pass vs fail judgements in cardinal proficiency research. Likewise, it provides the foundations for an argument towards a more gradual development of this important skill.

Strategy Assessment

This method of analysis is used infrequently when exploring cardinal proficiency development. Across the wider field of developmental psychology, the use of this method is scarce. This could stem from an historical focus on descriptions of children's behaviour at start and end points, or on the earliest manifestations of particular skills (Adolph et al., 2008). This focus on static states can mean the *process* of developmental change can be overlooked. Understanding *how* children acquire a particular skill requires detailed descriptions of behaviours during the period of change. For example, strategy assessment during the period of change establishes how children engage with a task. Looking at the strategies they use during the period of change can provide insight into how their knowledge and understanding is developing (Siegler,1996; Thelan, 2005). Strategy assessment does require regular sampling at intervals appropriate to the rate of change, and this can be time-consuming. However, the additional detail this methodology can offer is vast. Such methodology can help to visualise the shape of developmental change (e.g., linear vs step-like), assess the rate of change, and help to formulate and test theories of development (Adolph et al., 2008).

To our knowledge strategy assessment has only been utilised once when exploring cardinal proficiency development. This research adopted a very coarse measurement with only two testing sessions one week apart. Likewise, the results only identify two strategies 'counting' vs 'grabbing', suggesting that non-cardinal principle knowers grab and cardinal principle knowers count (Chetland & Fluck, 2007). Two sampling sessions are clearly not enough when accounting for the rate of change, as the current knower-levels theory suggests cardinal principle development happens over a number of months (Lee & Sarnecka, 2010, 2011; Litowski et al., 2020; Sarnecka & Carey; 2006; Sarnecka & Lee, 2009; Wynn, 1990). Correct implementation of strategy assessment would allow testing of this theoretical premise. Additionally, it would allow us to examine if children do gain knowledge of the cardinal principle when they are able to give five or more items, another premises within the knower-levels theory. If children begin to employ a new, more advanced, strategy when they are able to give five items, this may reflect an increase in their conceptual understanding and would therefore correspond with the knower-levels account of cardinal proficiency development. However, if this is not the case then maybe cardinal principle development is a more gradual change, as suggested by recent evidence (Krajcsi & Reynvoet, 2023).

Research Scope

The scope of this research responds to the theoretical and methodological problems raised above, and the emerging evidence suggesting a revision of the knower-levels account of cardinal proficiency development. The thesis comprises a meta-analytic assessment of the reliability of the Give-N task, a subsequent empirical assessment of these findings, and a micro-genetic analysis of children's strategies and performance whilst participating in cardinal principle tasks. The aim of this research project is to assess the current measures of cardinal proficiency development and consider if current theoretical accounts of cardinal proficiency are suitable. Specifically, we aim to answer the following research questions.

Research Questions

1. Is the Give -N task, the primary measure of cardinal proficiency, a reliable and robust measure?
2. As suggested by the knower-levels account of cardinal proficiency development, do we see conceptual advancement in cardinal proficiency when children are able to give five items?
3. Can other measures of cardinal proficiency help to validate or advance the knower-levels account of cardinal principle development?

Responding to these research questions will assess a key measure of cardinal principle development. Such assessment seems timely, as the reliability of this task has not yet been established. This should inform research practice, facilitating standardisation and optimal methods. Additionally, the research may support the utilisation of alternative methods to advance current practice and theory. Importantly, our research will be an initial step questioning the reliance on the Give-N task, and the critical interplay between current theory and the use of this task over the last thirty years. Study one finds key differences in the ways the Give-N task is implemented, and the effects of this were tested empirically within study two. Study three then explores the dynamic of change, and alternative measures of cardinal proficiency. Each study within this alternative format thesis will appear in order and in publication, with short summaries connecting each study. The final section will summarise and discuss the overall findings of this research project.

Experimental Chapter 1: Examining the Give-N Task as a Measure of Cardinal Proficiency

One of the first key milestones in young children's mathematical development is understanding the cardinal principle. It is at this point children understand that the final number in the count represents the number of items in the set (Chu et al., 2016; Geary et al., 2018). Understanding the cardinal principle is the gateway to more complex mathematical operations, such as addition, subtraction and multiplication. It is therefore imperative children have a secure understanding of this principle. The earlier children acquire the cardinal principle, the earlier they can progress onto more complex mathematics, and this advantage is maintained throughout their education (Aunio & Niemivirta, 2010; Chu et al., 2015; Chu et al., 2016; Jordon et al., 2009; Koponen et al., 2013). Therefore, understanding how this skill develops, and how we can scaffold children's learning during this period is essential. Equally, it is essential that research responding to this is reliable, and the measures used to assess cardinal principle acquisition are fit for purpose.

Within the literature exploring cardinal principle development and young children's early numerical skills the Give-N task is the primary measure. This measure is used ubiquitously and is assumed to be reliable. Research assertions and theoretical stances are built upon results from this measure. The reliability and validity of this task needs to be secure. This paper focuses on this and assesses the reliability of the Give-N task as a measure of cardinal proficiency.

This chapter contains the first paper of this thesis which responds to the first research question: *Is the Give -N task, the primary measure of cardinal proficiency, a reliable and robust measure?*

This paper is currently in preparation for publication:

Thomas, H., Lewis, C., & Dunn, K. (in preparation). A Systematic Review and Meta-analysis of the Give-N task as a Measure of Cardinal Proficiency.

A Systematic Review and Meta-analysis

of the Give-N task as a Measure of Cardinal Proficiency.

Hannah Thomas

Charlie Lewis

Kirsty Dunn

Contributions

HT, KD and CL designed the study, with HT and CL writing the manuscript. HT undertook the full systematic review and meta-analysis, including all statistical analyses. CL performed the secondary review process.

Abstract

Development of the cardinal principle is a key milestone in children's understanding of number, here children understand the final number in a count represents the total number of items. The Give-a-Number or Give-N task is the primary measure when assessing cardinal proficiency. To date the robustness of this task has not been assessed systematically. Following PRISMA and NIRO guidelines we analysed 118 studies utilising this method. Presentation of task and the effects of this were explored. We found that the Give-N task is scored and implemented in many different ways, and in turn this affected children's performance. Specifically, how children were asked for the numbers of items, if the sets of items were changed during the task, and presenting the task during pretend play influenced children's displays of cardinal proficiency. Importantly, these influences are not accounted for within current theoretical accounts of cardinal principle development.

Introduction

Cardinal proficiency, understanding that the final number in a count represents the total number of items, is an important foundation for future mathematical achievement (Chu et al., 2016; Geary et al., 2018). The relationship between early mathematical skills, such as cardinal proficiency, and improved outcomes in informal arithmetic, calculations and mathematical reasoning has been recognised (Chu et al., 2016). Children who develop these skills earlier have higher mathematical competence and maintain this advantage throughout their education (Aunio & Niemivirta, 2010; Chu et al., 2015; Chu et al., 2016; Jordon et al., 2009; Koponen et al., 2013). This mathematical advantage has individual and societal importance, as people with higher mathematical competence are more likely to be in employment, have improved earning potential, and experience superior health outcomes (Brynner, 2004; Reyna et al., 2009). As cardinal proficiency is a vital skill for later mathematical competence, and subsequently improved life outcomes, it is imperative that we are able to measure this competency accurately and consistently. It is important, therefore, to reassess the reliability of the Give-N task as a measure of cardinal proficiency.

The Give-N task is a primary measure of young children's cardinal proficiency (Crooks & Alibali, 2014), and is used universally throughout mathematical and wider developmental literature. This task assesses a crucial milestone in children's arithmetical development and has been utilised in a number of influential studies, and guided prominent theoretical positions. To date there has been no review of the Give-N task, and no assessment of the task's consistency. It is important, therefore, to examine whether the Give-N task is an accurate and reliable measure of cardinal proficiency. By undertaking a systematic review and meta-analysis we respond to these issues.

The Give-N task as an instrumental and ubiquitous measure of cardinal proficiency

With cardinal proficiency being such an important developmental milestone for preschool children this skill has been explored and associated with many other competencies. For example, language and working memory abilities (Purpura & Ganley, 2014), sharing capabilities (Chernyak et al., 2016; Chernyak et al., 2019), selective attention (Brueggemann & Gable, 2018), problem solving (Chu et al., 2018), and spatial skills (Verdine et al., 2014). Crucially, the Give-N task was utilised across all these research fields, demonstrating its ubiquitous use throughout the developmental literature.

In addition, the Give-N task has occupied a prominent position within intensely debated areas of developmental psychology. This includes the discontinuity vs core knowledge debate in the acquisition of the counting principles, wherein the diverging positions dispute the innate foundations of the counting principles (Gelman & Gallistel, 1978; Le Corre et al., 2006; Sarnecka & Lee, 2009; Sarnecka & Carey, 2008; Wynn, 1990, 1992). How the development of these skills is perceived is important, as this affects how these skills are nurtured and encouraged through parental and educational means. In addition, the Give-N task features heavily in the debate regarding the utility of the Parallel Individuation System, the Approximate Number System and the Analogue Magnitude System to later arithmetical development (Dehaene, 2011; Hyde et al., 2017; Starr et al., 2013; Wagner & Johnson, 2011; Wilkey & Ansari, 2020). Again, the importance of such exchanges is significant, with the products informing how numerical skills are facilitated. Notably, irrespective of stance, the Give-N task pervades the literature. These debates continue today, with advancements in methodology and technology used to explore these issues. However, the Give-N task remains the one constant, maintaining its instrumental role within the field.

Furthermore, the Give-N task has been, and remains, influential in current theories of cardinal proficiency development, in particular the knower-level account (Sarnecka & Lee, 2009; Lee & Sarnecka, 2010, 2011; Wynn, 1990). The Give-N task and the knower-level account are inextricably linked, with findings from the task substantiating the knower-level account of cardinal development. This account of cardinal proficiency development has added, and continues to accumulate, a substantial evidence base with increasing levels of statistical sophistication. However, this empirical evidence remains grounded in results from the Give-N task, signifying the influential nature of the measure.

Having established the ubiquitous and influential nature of the Give-N task, it is therefore essential to examine the consistency of how this task is presented to children, and if differences in presentation affect children's ability to pass the task. If children's cardinal proficiency is affected it introduces an additional element of variation across studies, making the results less comparable and limiting theoretical conclusions across many research domains. Furthermore, this may impact upon highly debated issues concerning children's mathematical development, and raise questions about prominent accounts of cardinal proficiency development.

Differences in presentation of the Give-N task – does this matter?

Typically, the Give-N task requires children to give or produce a sub-set of items from a larger set. Children are deemed to have passed the task, and are credited with cardinal proficiency when they are able to give five or six (or more) items. An initial scoping search of research articles utilising the Give-N task revealed an acute lack of systematicity in how this task is presented to children, and how children are assigned cardinal principle status during the task. In a number of studies children are asked to give items to the experimenter whereas in others they are asked to give items to a puppet, with some providing a reason or story for why the child should give the items, and others not. Sometimes the items children are asked to give changes during the task, and they are asked if they have given the correct amount, and in other studies not. Likewise, the number of items presented to the child from which to make a sub-set varies greatly between studies; ranging from anywhere between ten to thirty items. How children are asked to give different numbers of items also varies, with several studies adopting a titration method (i.e., asking for N plus one if the child is able to give N, or N minus one should they not be able to give N), others requesting the number of items incrementally, and still others requesting items in a random numerical order.

Remarkably, these are only the differences in presentation of the Give-N task. How children are classified as cardinal principle knowers also varies extensively between studies. Some studies assign cardinal principle status when children can give five items, others when children can give six items. Likewise, some adopted a ‘two out of three correct’ criterion when assessing cardinal proficiency (i.e., if children receive three trials for ‘give five’ they would need to correctly give five on two out of the three trials), whereas others required children to give the correct number on all trials. A multitude of differences were also noted with regards children’s errors on the task and how this affected assessment of their cardinal proficiency. These include scoring criteria such as not giving ‘N’ when asked for a different number (e.g., if a child gave six when asked for four this would be counted as an error against the number six and the number four), with some studies allowing children one error based on this criterion, others allowing no errors, and a final group not adopting this criterion at all. Furthermore, these different criteria, when assigning cardinal principle status, are combined and utilised in numerous ways.

In addition, and of note, a small number of studies used a distinctly different procedure, adopting a role play methodology. Whilst these studies did not discuss why this approach was used, it is important to examine the effects of this different procedure. Previous research has highlighted the importance of context during mathematical and problem-solving tasks (Donaldson, 1978; Nunes Carraher et al., 1985), with different outcomes seen between naturalistic situations and school-based, context-reduced routines. Specifically, children performed better when the mathematical problems were presented in a natural context, typical of their daily lives (Nunes Carraher et al., 1985). The types of role play adopted by some studies in our scoping search reflect real life play situations that children participate in frequently, such as playing shops. It may be that such a methodology recruits different routines and facilitates children's cardinal proficiency. Importantly, the effect of this methodology needs to be measured, and potentially incorporated into current theories of cardinal development.

Critically, the scoping search has highlighted vast inconsistencies in the way in which the Give-N task is implemented, and these inconsistencies need to be measured to allow assessment of this task as an accurate and reliable measure of cardinal proficiency. The differences described may introduce variation in children's pass rates. For example, the use of a check question (asking the child if they have given the correct number of items e.g., 'is that five?'), could allow for reflection and decomposition of the task, refreshing / refocusing working memory and orienting children's attention to the relevant aspects of the task (Andrews et al., 2003; Barrouillet et al., 2009; Bertrand & Camos, 2013; Halford et al., 1998), thus reducing errors. Likewise, changing the sets of items children were asked to give during the task (i.e., children were asked to give four apples, then five cars), may promote sustained attention and therefore reduce the number of errors children make (Brueggemann & Gable, 2018). Again, the effects of these inconsistencies need to be measured, and potentially incorporated into current theories of cardinal development.

The prominent theoretical positions regarding cardinal proficiency take opposing views of how this important skill develops. The continuity stance posits that innate counting principles underpin symbolic number development, and it is simply superior counting procedures that allow acquisition of the cardinal principle (Gallistel & Gelman, 1992; Gelman 1993; Gelman & Meck, 1983). Whereas, the discontinuity position suggests that acquisition of the cardinal principle represents a conceptual change. Proposing that innate representational systems, such as parallel individuation, provide the foundations for cardinal development, and bootstrapping processes integrate this system with other representational systems such as natural language. This gives rise to a new representational system, and the concept of symbolic number (Carey, 2004; Lee & Sarnecka, 2010). Whilst these accounts provide diverging accounts of cardinal development, both positions agree on the importance of cardinal proficiency to children's mathematical development. Crucially, however, neither theoretical account acknowledges how the presentation of a task the child is asked to perform may influence their displays of cardinal proficiency. With the lack of systematicity found in our scoping search, coupled with both theoretical stances utilising the Give-N task ubiquitously to progress and refine their positions, it is essential the effects of presentation are analysed and reflected upon when advancing current theories of cardinal principle development.

Purpose of the current study

As discussed, cardinal proficiency is an important accomplishment in preschool children's mathematical development, and plays a significant role in their future mathematical achievement. This is recognised empirically and theoretically. Additionally, this key skill has been explored and associated with many other abilities, encompassing a multitude of research domains. Crucially, use of the Give-N task to measure cardinal proficiency is ubiquitous within the literature. Given the extant amount of research using this task, it is essential to assess the methodological consistency across studies, and measure the effects any inconsistencies may have on children's successful achievement of this task.

Reflecting on the above, we aim to reveal the extent to which the presentation of the Give-N task varies, and to measure the effect these differences may have on children's accomplishment of the task. Specifically, we answer the questions: What are the different ways the Give-N task is presented to children? Regardless of the method of presentation, does the Give-N task produce consistent results? If not, do the differences in presentation account for the variance in the results? By undertaking a systematic review and meta-analysis we will provide empirical evidence in response to these issues. This should help to standardise the procedure for what has become a critical test when assessing children's cardinal proficiency, and would enable improved comparisons across a vast number of studies and numerous fields of research. Furthermore, this review may advance theoretical positions by introducing the effects of natural contexts and motivational features into the debate of how cardinal proficiency develops.

Method

The Preferred Reporting Items for Systematic Review and Meta-analysis (PRISMA-P, Moher et al., 2009; Moher et al., 2015; Page et al., 2021) and Non-Interventional, Reproducible, and Open (NIRO, Topor et al., 2020) guidelines were followed for this review.

Research protocol

A research protocol was formulated and pre-registered on the 15 July 2020, and can be accessed via the Open Science Framework (https://osf.io/tjbph/). No deviations from this were made.

Search strategy

A systematic search strategy was developed in collaboration with the University Faculty of Science and Technology librarian. Title words, key words, and author key words were extracted from ten key papers, and where possible thesaurus terms were used to improve the sensitivity of the search. See Figure 1 for exact search string. Search fields included the Title, Abstract, Keywords. A general search, without reference to the Give-N task, was also carried out for a secondary meta-analysis incorporating other measures of cardinal proficiency. The searches were performed on the 11 July 2020 and 6th November 2020 respectively.

("cardinality" OR "cardinal* principle" OR "cardinal* knowledge") OR ("give-x" OR "give x" OR "GaN Task" OR "give-a-number" OR "give n" OR "give-n") AND ("numeracy*" OR "numeracy skill*" OR "numeracy abilit*" OR "math* abilit*" OR "math*" OR "math* achievement") AND (preschool* OR pre-k* OR pre-school OR kindergarten* OR nurser*)

Figure 1
Exact search strategy string used for all databases

Sources

The Psycinfo, ERIC, CINAHAL and Scopus databases were selected for the search. The references of all eligible papers were also examined to ensure maximal scope. To access grey literature ProQuest Dissertations and Theses was searched. Additionally, authors were contacted when information was required for inclusion of eligible studies. Any unpublished data was requested at this stage.

Inaccessible papers

Following the initial search 12 papers were identified as inaccessible. A University Librarian was able to source six of these papers, and the small number of unobtainable papers are unlikely to have affected the results.

Screening

Software

Zotero reference manager was utilised for the screening process, facilitating a more transparent process when removing duplicates (Staaks, 2020). Rayyan was used for full text reviews (Ouzzani et al., 2016).

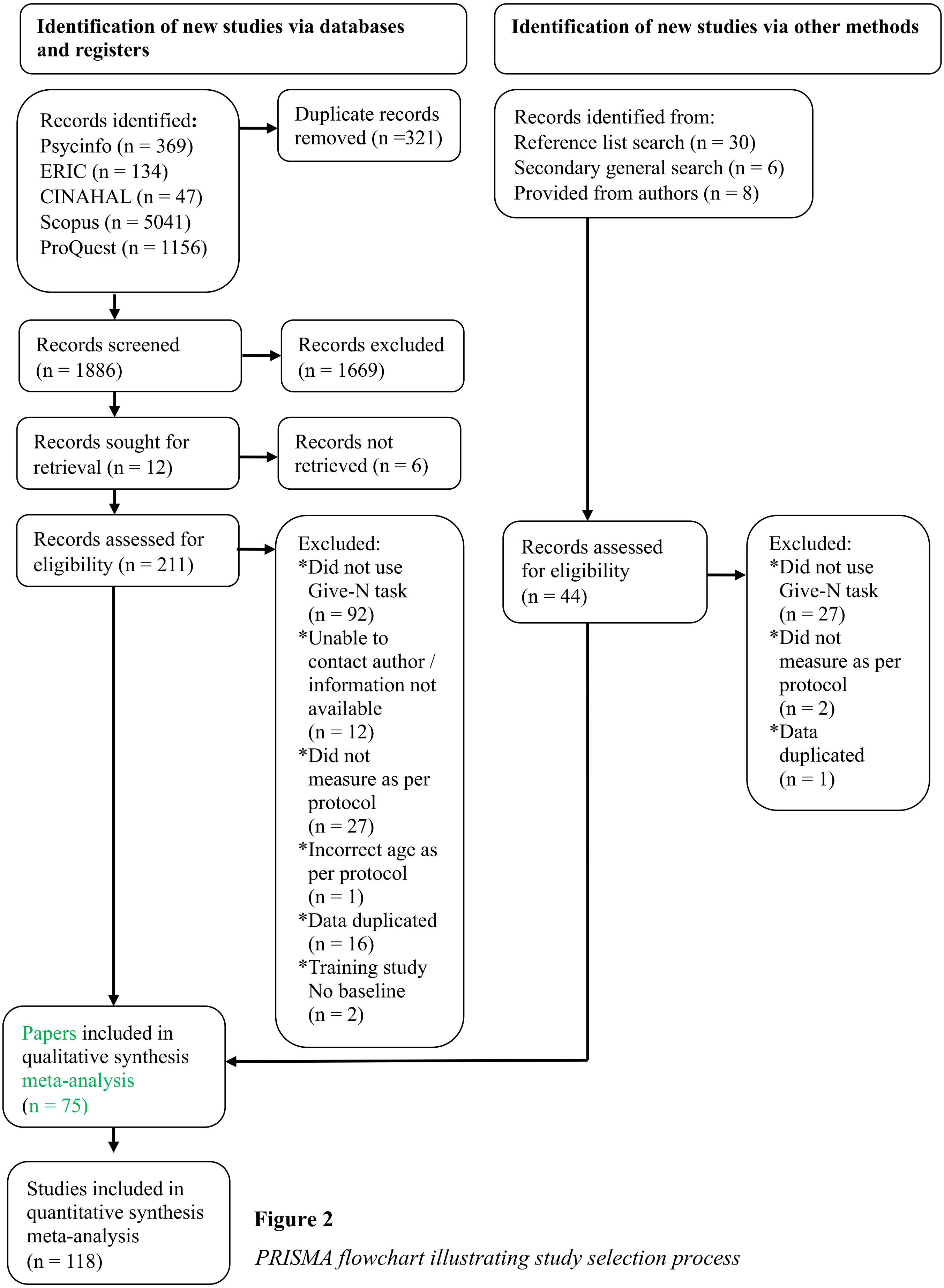

Figure 2

PRISMA flowchart illustrating study selection process

Process

Initial title and abstract screening was undertaken by one reviewer. Full text reviews were carried out by two independent reviewers, with the second reviewer assigned a random sample of 27% (see Table 1 for inclusion and exclusion criteria). The selection process is summarised in Figure 2.

Data extraction

The following information was extracted from the papers: Participant age / age groups within each study (incl M & SD), number of participants (incl number of boys / girls if available), participant SES and ethnicity if available, proportion / percentage pass rates, method / description for Give-N task, date of study, place / country where study was conducted.

Data extraction was carried out by one reviewer, with a second independent reviewer undertaking data extraction for a random sample of 27% (reliability of 96% was observed, with any differences resolved by revisiting the research paper). Nineteen papers did not provide explicit proportions, percentages or numbers of children deemed cardinal principle knowers. These authors were contacted via email to obtain the information. Nine authors provided information for 20 studies, of which 18 were included in the meta-analysis. Seven authors did not reply, and three were not able to obtain the information required. These studies were therefore excluded.

Table 1

Inclusion and exclusion criteria for eligible studies

	Inclusion criteria	Exclusion criteria
Study design	Observational	Review papers
	Cross-sectional	Theoretical papers
	Experimental [a]	Case studies
	Intervention [b]	

	Inclusion criteria	Exclusion criteria
Participants	Typically developing children Aged 2 to 5 years Attending mainstream public or private pre-schools, schools or kindergartens	Children with diagnosed learning difficulties or developmental disorders Children with any mathematical difficulties Children attending specialist education establishments
Time	No restrictions	No restrictions
Language	Full texts will be in English or Where possible other languages will be included dependent on translation provision	None specified
Intervention	Cardinal principle acquisition must be measured using the Give-N task in any form	None specified
Outcome	The proportion, percentage or number of children deemed cardinal principle knowers, as measured by the Give-N task is required	None specified
Setting	Mainstream pre-school Mainstream school Kindergarten Research laboratories	None specified

Results

In total, 118 studies using the Give-N task were available from the 75 publications included in the review (see https://osf.io/tjbph/ for summary).

Risk of bias and study quality

The majority of studies were assessed as low risk of sampling, selection, performance, outcome, measurement bias and reporting bias. See https://osf.io/tjbph/ for risk of bias spreadsheet and written summary.

Publication bias

The first analyses were conducted to assess risk of publication bias within this dataset. No asymmetry was identified on either funnel plot (see Figure 3), and Egger's regression tests between effect size and standard error (z = -1.42, p = .16) and effect size and sample size (z = 1.71, p = .09) were nonsignificant. We conclude that risk of bias was low. Closer inspection of the funnel plots revealed high levels of variance in children's performance across the studies, with a large number falling outside the 95% CI. Given that a large proportion of the studies had small sample sizes, this suggests that a small sample bias may be present.

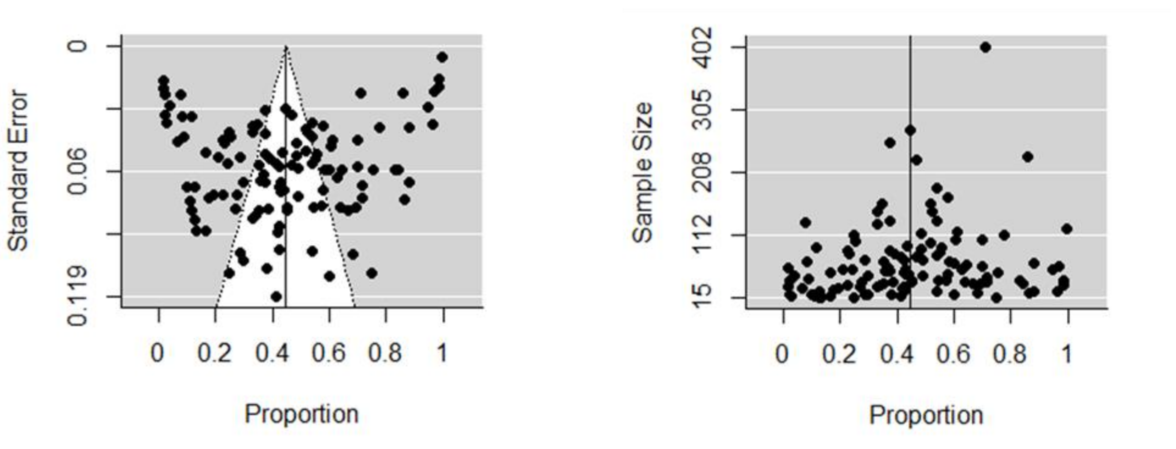

Figure 3

Funnel plots of individual effect sizes plotted against standard error and sample size

Meta-analysis

To examine the overall variance in children's pass rates for the Give-N task, a meta-analysis was performed in R using the Metafor and Meta package version 4.0.3. To assess if the use of raw proportional data influenced the analyses, effect sizes were calculated in three different ways; using the raw proportions, a logit transformation, and a double-arcsine transformation (Wang, 2018). All three effect size calculations produced the same summary effect size. The raw proportional data was therefore used throughout the analyses.

Main model

To assess the consistency of children's pass rates a standard meta-analysis was undertaken comprising all study effects. A random effects model was employed, as the studies included in the analysis are unlikely to be functionally or characteristically equivalent, so it is appropriate to allow the true effects to vary between studies (Harrer et al., 2019; Wang, 2018). Allowing the effect sizes to vary between studies was also essential to our research questions. The Paule and Mandel Tau^2 (PM) estimator was utilised as this is suggested to be more robust and is thought to have superior performance compared to other commonly used estimators, such as the DerSimonian-Laird method (Novianti et al., 2014; Veroniki, 2016). Likewise, the PM estimator is recommended when the between-study variance may be large, as suggested by the funnel plot (Novianti et al., 2014; Veroniki, 2016).

The overall random-effects model summary proportion was .45, 95% CI [.40, .49] (see Figure 3 for forest plot). This suggests that on average 45% of children aged two – five (inclusive) are able to pass the Give-N task by giving five or six items, and are therefore classified as cardinal principle knowers. The model revealed a high I^2 value ($I^2 = 97.76$) suggesting a large amount of the variance in children's pass rates is attributable to true between study heterogeneity, not sampling error (Higgins et al., 2003). Using the Predict function in the Metafor package, it was estimated that in 95% of populations the pass rates could fall between 0% and 94%. See Table 2 for full model statistics.

Table 2

Meta-analytic results of the main model

	Random-effects model				
Outcome measure	*k*	$\bar{p}$	*p* value	I^2	τ^2
Pass rates	118	.45 [.40, .49]	<.001	97.76 [97.13, 98.30]	.06 [.05, .08]

Influential cases

A Baujat plot was constructed to identify any influential studies (see Figure 5). This suggested that six studies (54, 18, 66, 100, 55 & 67), largely with older children, may have had a greater influence on the results. However, more specific 'leave one out analyses' suggested none of the studies exerted a significant influence on the results, so no studies were removed.

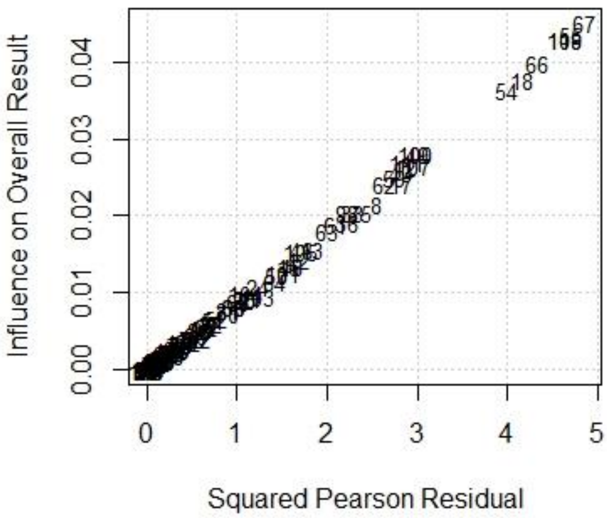

Figure 5

Baujat plot of influential studies

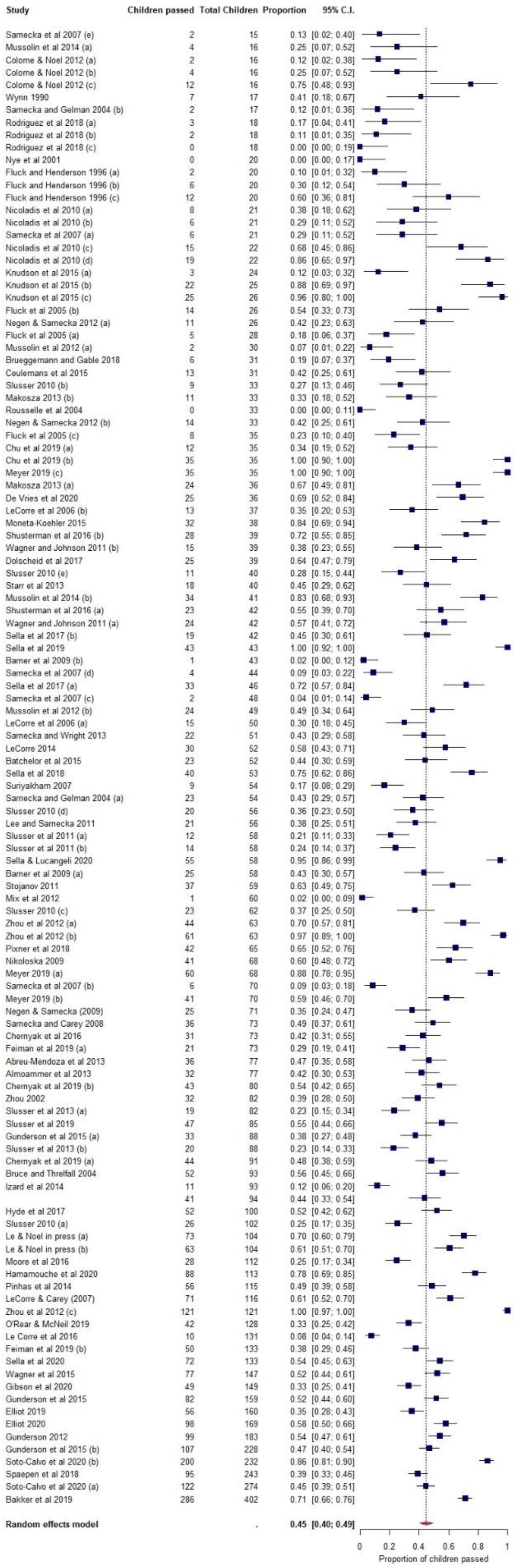

Figure 4

Forest plot depicting a summary of all study effect sizes

Moderator analysis

Given the large amount of unexplained between-study heterogeneity, further examination of potential moderators was necessary. Reflecting on the findings of our scoping search, the various different routines employed during the Give-N task were explored. We identified seven differences in presentation of the Give-N task (see Table 3), and 18 different ways children were assigned cardinal principle status (see Table 4). To assess if these methodological differences could explain any of the between-study variance meta-regression analyses were undertaken.

Table 3

Description of presentation methods and their proportional use across studies

Implementation difference	Description	Proportional use		
		Toy	Experimenter	Table
Who the child was asked to give items to	Child was asked to give items to a toy/puppet, the experimenter, or put them on the table/box	.73	.18	.09
		Incremental	Random	Titration
How the child was asked for the items	Child could be asked for different numbers of items in an incremental or random order, or by titration method	.09	.39	.52
			Yes	No
Was a check question utilised	Was the child asked if that was the correct number of items		.63	.37
		Small (under 15)	Medium (15-19)	Large (20+)
Number of items presented to child	The number of items presented to the child from which they made a sub-set	.30	.56	.14
			Yes	No
Change of items	Did the type of items the children were asked to give change during the procedure (e.g., asking them to give apples, then asking for toy cars)		.29	.71

Implementation difference	Description	Proportional use		
			Five	Six
Five vs six	Was the child classed as a cardinal principle knower if they could give five OR six		.63	.37
		1 (no reasoning)	2 (general reason)	3 (pretend / role play)
Engagement level	The reasoning / story / role play presented to the children when asking for items	.77	.14	.09

Note. Some studies did not report the number of items presented to the child or used a varied number of items (*n* = 29), who the child gave the item to (*n* = 3), the way in which they requested the items - titration / random / incremental (*n* = 2), if a check question was utilised (*n* = 2), if the items were changed during the procedure (*n* = 2), if children were cardinal principle knowers when they could give five or six (*n* = 2), or how the task was presented to the child (*n* = 2).

Table 4

Description of how children were assigned cardinal principal status when asked to give five or six, and their proportional use across studies

Method of assigning cardinal principle status on give five / six trials	Proportional use
Correct 2 out of 3 times (no stipulation on errors in previous sequence / giving N when asked for a different number)	.24
Correct once only asked once (no stipulation on errors)	.13
Correct all trials (no stipulation on errors in previous sequence or giving N when asked for a different number)	.07
Correct all trials only 1 error allowed in previous sequence	.01
Correct once out of three attempts	.02
Correct all trials - no errors allowed (incl not giving that numerosity for another number word)	.01
Highest number given – no errors allowed (incl not giving that numerosity for another number word)	.01
Correct once only asked once (and allowed 1 error when giving 5 for another number)	.01
Highest number given (1 error allowed in previous sequence)	.03
The task was stopped if the child failed on consecutive attempts.	.03

Method of assigning cardinal principle status on give five / six trials	Proportional use
Bayesian method	.06
Pre-Bayesian 67% successes (with a minimum of two trials) at a given number N , and at least 67% failures (with a minimum of two trials) at N + 1. Failures were counted against both numbers involved. For example, if a child gave four apples when asked for "two," that was counted as failure on both "two" and "four"	.03
Give the correct number (N) 2 out of 3 trials, and of those times that the child provided (N) two-thirds of the times the child did so it was in response to a request for (N).	.02
Give the correct number (N) 2 out of 3 times, and did not give (N) more than once for any other trials	.09
Give the correct number (N) 2 out of 3 times and did not give (N) for any other trials	.01
Give the correct number (N) 2 out of 3 times, and had twice as many successes as failures on trials for 5 & 6	.05
Give the correct number (N) 2 out of 3 times, and not give (N) when asked for another number at least 2 out of 3 trials (only gave N once out of 3 trials when asked for another number)	.02
Give the correct number (N) 2 out of 3 times, and not give n when asked for another number more than half as often (percentage-wise) when asked for other set sizes than when asked for (N) itself. For example, a child who gave two items 80% of the time when asked for two, was scored consistently correct on two only if he or she gave two items no more than 40% of the time when asked for one, three, five, and six items	.16

Note. Two studies did not report the method.

Meta-regression

Adopting a maximal to minimal approach, children's age was entered into the model as a covariate, as older children are more likely to be cardinal principle knowers (Litkowski et al., 2020). Simultaneously, the different methods of presentation were entered as categorical moderators. The number of items presented to the child was not entered as a predictor because 23 studies did not provide this information within their methods. The different scoring methods were also not included as these, with under 10 studies per method, had an insufficient distribution for moderator variables (Higgins, 2016).. Additionally, seven studies were excluded from the analysis (Hyde et al., 2016, Fluck et al., 2005$_{(c)}$, Wagner et al., 2015, Feiman et al., 2019 $_{(a) \& (b)}$, Chu et al., 2019 $_{(a) \& (b)}$) as they did not provide enough detail to code the moderator variables.

Results

The meta-regression was significant $Q(10) = 256.92$, $k = 111$, $p < .001$, $I^2 = 89.78$, with the model explaining 72.09 % of the between study variance in children's pass rates ($R^2 = 72.09\%$). As expected, age was a significant moderator of children's pass rates, as were three additional moderators. Not changing the set of items and asking for the items in a random order during the Give-N task reduced children's pass rates. Engaging children in pretend play while conducting the test increased their pass rates. See Table 5 for full model statistics.

To assess the individual contribution of the significant moderators, pseudo R^2 values were calculated. As expected, age accounted for most of the between study variance (pseudo $R^2 = 63.01\%$). How the child is asked for the items explained just over one sixth of the between study variance (pseudo $R^2 = 16.70\%$), and level of engagement and change of items each accounted for a similar amount of variance; approximately 10% (pseudo $R^2 = 9.64\%$ & pseudo $R^2 = 10.05\%$ respectively).

Table 5

Meta-regression model including age and implementation methods as moderators

	Random-effects				
Outcome measure	*K*	*Q*	*p* value	I^2	τ^2
Pass rates	111	256.92	<.001	89.78 [86.07, 92.24]	.02 [.01, .02]
	Moderators				
	B	SE	Z	*p*	
Intercept	-0.55	0.10	-5.36	<.001	
Age	0.25	0.002	11.96	<.001	
Exp. vs toy (table)	0.10	0.06	1.55	.12	
Exp. vs toy (exp.)	-0.10	0.04	-0.30	.77	
Check quest. (no)	-0.02	0.04	-0.54	.59	
Titration vs random (random)	-0.17	0.04	-4.14	<.001	
Titration vs random (incremental)	-0.09	0.05	-1.67	.10	
Change of items (no)	-0.12	0.04	-3.08	.002	
Five vs six (give 5)	-0.03	0.03	-0.77	.44	
Engagement (reasoning)	-0.03	0.04	-0.69	.49	
Engagement (pretend play)	0.19	0.07	2.66	.007	

Note. Where mean age was not available ($n = 9$) the median was used. Redundant moderators were automatically dropped from the model.

With only four of the variables indicated as significant moderators, a LogLik Likelihood Ratio test comparing the full and a reduced model revealed that reducing the model to only include significant moderators did not improve the model fit. Additionally, a lower AIC suggested the fuller model was a superior fit (see Table 6 for model comparisons).

Table 6

Model comparisons for full model, with age and implementation methods as moderators, and reduced model with age, change of items, how the child was asked for the item, and level of engagement as moderators

Model name	Nested model	Model fit				LRT test against nested	
		AIC	BIC	LL	Df	Df	χ^2
Full model	-	-101.14	-68.62	62.57	12		
Full model	Reduced model	-105.71	-84.04	60.86	8	8	3.42

Discussion

The current meta-analysis has highlighted the many different ways in which the Give-N task is implemented and scored. Within 118 studies (from 75 publications) there were seven different ways in which the Give-N task is implemented and 18 different ways the Give-N task was scored. In addition, our findings suggest that children's pass rates on this task are highly variable. Moderating variables (in order of significance) were age, asking for the items in a random numerical order, not changing the sets of items in-between trials, and utilising a pretend play methodology. Using pretend play methodology and increase in age increased children's pass rates, whereas asking for items in a random order and not changing the sets of items decreased children's pass rates.

This review and meta-analysis has answered all three of our research questions. We have revealed that, as expected, the Give-N task is implemented in many different ways, and these differences in presentation do affect children's ability to pass this task. Given these variations, the results suggest that the Give-N task may not be a reliable measure of children's cardinal proficiency and, given the ubiquitous nature of this task, this is potentially problematic. Importantly, our results are able to account for the majority of the variance found in children's pass rates. The primary explanation for the variance is age; the older the child the more likely they are to pass the task. This is to be expected, as older children are typically more likely at the age when the development of cardinality is acquired. We will examine this result and the remaining moderator variables; random numerical order when asking for items, not changing the sets of items, and pretend play next.

Moderators

Asking for Numbers of Items in Random Order

Asking for the items in a random numerical order reduced children's ability to pass the task. The Give-N procedure is already a complex task and imposes a significant load on children's working memory (Halford et al., 1998). Children must hold in mind the number of items requested whilst they count out or pass the items, with higher numbers of requested items extending the length of these trials therefore making these more demanding. With this in mind asking for numbers of items randomly, incongruent to the child's count list, may introduce an element of switching (Wright & Diamond, 2014), thus making the task more difficult. It is understandable that asking children for two items, then six items may be more confusing, jumping from smaller numbers to larger numbers or vice-versa reduces regularity and predictability which may hinder children's displays of cardinal proficiency. Likewise, the inconsistency in the numbers requested may obscure children's understanding of the task, making them less likely to engage and attend to the task, therefore reducing their displays of cardinal proficiency.

Changing the Sets of Items

Not changing the sets of items in-between requests also reduces children's pass rates. Some researchers used only one set of items, for example 15 small toy pigs on each trial, and used these throughout the task. Whereas others had three or more sets of items e.g., 15 pigs, 15 ducks and 15 frogs, and after requesting three pigs the following trial requested four ducks. Thus, changing the sets of items throughout the task. Children's displays of cardinal proficiency were reduced in the studies where researchers did not change the sets of items. Once again, it is likely that this is due to engagement, not changing the sets of items makes the task more tedious and children's sustained attention wanes (Brueggemann & Gable, 2018), thus making them more error prone and limiting their displays of cardinal proficiency.

Pretend Play

A number of researchers incorporated an element of pretend play whilst administering the Give-N task, and this facilitated children's displays of cardinal proficiency. Whilst this wasn't an unexpected result, as this alternative methodology was found in our scoping search, it wasn't clear if this would have an effect on children's displays of cardinal proficiency. Our results suggest that, again, engaging children and sustaining their attention during this cognitively demanding task is making them less error prone and facilitating their cardinal proficiency capabilities. However, with this particularly divergent method it is not clear if it is the naturalistic context that engages children and facilitates their ability (Donaldson, 1978; Nunes Carraher et al., 1985), or if this effect is specific to pretend play routines. It may be that engaging in pretend play activities frees children from the 'rules' of reality allowing them to reason about number in different ways (Diaz & Harris, 1990; Weisberg & Gopnik, 2013).

Conclusion

Our findings have implications for practice and theory. Practically, we highlight the need for the Give-N task to be standardised. This task is a ubiquitous, primary measure of cardinal proficiency development and is utilised throughout the developmental literature (Brueggemann & Gable, 2018; Chernyak et al., 2016; Chernyak et al., 2019; Chu et al., 2018; Crooks & Alibali, 2014; Purpura & Ganley, 2014; Verdine et al., 2014). It is therefore important this task is presented and scored in the same way as we now know that differences in presentation can affect children's displays of cardinal proficiency. Not only will this facilitate easier comparisons across studies, but it will also enhance professional standards within this field of research. Our findings also suggest that other measures of cardinal proficiency should be considered. The Give-N task dominates this field of research, and the results here suggest that this task may not be robust and may produce inconsistent results. Whilst standardisation of this task will improve this, it is important to consider the use of other measures to triangulate results when using this task. Consequently, increasing the reliability and validity of cardinal proficiency research.

In addition, we highlight the need for current theoretical positions regarding cardinal proficiency development to be reconsidered. The dominant, knower-levels, account (Sarnecka & Lee, 2009; Lee & Sarnecka, 2010, 2011; Wynn, 1990) of how this skill develops does not acknowledge the possible effects of task presentation. It is important that current accounts of cardinal proficiency development incorporate the possible effects of natural contexts and engagement when asserting how this crucial skill develops. Furthermore, we highlight the need for the knower-levels account of cardinal proficiency development to be re-visited. This account and the Give-N task are inextricably linked. The task forms the basis of the theoretical description, and the assumptions of the theoretical account shape the constraints of the task. For example, numbers higher than five are typically not requested, as it is assumed that once a child can give five, they are cardinal principle knowers. Our meta-analytic results suggest that the Give-N task is not robust or reliable, therefore findings from this task that underpin the knower-levels account need to be re-assessed using other measures of cardinal proficiency to validate the theoretical assertions.

Re-visiting the knower-levels account corresponds with recent discussions suggesting that development of the cardinal principle may be gradual rather than, as suggest by the knower-levels account, a sudden conceptual change when children can give five items (Krajcsi & Fintor, 2023; Krajcsi & Reynvoet, 2023). This emerging evidence utilises alternative measures alongside the Give-N task to assess the central tenets of the knower-levels account. Whilst the findings partially support the knower-levels account, they do suggest that over-reliance on the Give-N task has limited this account of cardinal proficiency development, and revisions need to be made. Our findings support this, as over-reliance on an unreliable task is problematic for any theoretical account. Furthermore, as mentioned above, the knower-levels account does not acknowledge the effects of task presentation on children's displays of cardinal proficiency and therefore needs re-visiting.

Undertaking a systematic review and meta-analysis of this critical task has allowed us to assess the reliability in an unbiased, objective way. Following PRISMA and NIRO guidelines has ensured a disciplined and systematic approach, improving the quality and comprehensiveness of the review. Consequently, we were able to include 118 studies in the meta-analysis, which has ensured a diverse, broad range of data from many research domains have been included, improving the reliability and validity of the analysis. However, with any meta-analysis the comparison of studies is limited as potential unknown inconsistencies between research groups or between studies cannot be fully accounted for. Additionally, we were unable to include the scoring methods in our moderator analysis due to the large number of different methods, which limits the findings regarding these. Nevertheless, we have been able to highlight the inconsistencies in scoring, and we have been able to account for most of the variance between the studies without the inclusion of scoring methods. Finally, our finding that pretend play facilitates cardinal proficiency requires further analysis as many of the studies that adopted this methodology were from the same research group. Therefore, we cannot be sure if this is a direct effect of pretend play or a spurious effect of a particular researcher. Likewise, we need to determine if this result is a due to the effects of pretend play or their use of naturalistic contexts, as discussed previously. This requires further empirical examination.

Importantly, the review has revealed that the Give-N task is implemented in many different ways, and these differences do affect children's displays of cardinal proficiency, suggesting this task may not be reliable. As such, we need to prioritise standardisation of this ubiquitous task. Furthermore, we must re-visit the dominant theoretical accounts of cardinal proficiency development, as their over-reliance on an unreliable task raises questions about the key tenets proposed within their account. Likewise, these accounts need to acknowledge the effects of task presentation with their descriptions of how children develop the cardinal principle. Finally, the effects of pretend play on children's number skills requires further exploration. Are the effects of pretend play unique, or are the effects due to natural contexts or simply engagement?

References

Andrews, G., Halford, G.S., Bunch, K. M., Bowden, D., & Jones, T. (2003). Theory of mind and relational complexity. *Child Development, 74*(5), 1476-1499.

Aunio, P., & Niemivirta, M. (2010). Prediciting children's mathematcial performance in grade one by early numeracy. *Learning and Individual Differences, 20,* 427-435.

Barrouillet, P. N., Gavens, N., Vergaywe, E., Gaillard, V., & Camos, V. (2009). Working memory span development: A time-based resource-sharing model account. *Developmental Psychology, 45*(2), 477-490.

Bertrand, R., & Camos, V. (2015). The role of attention in preschoolers' working memory. *Cognitive Development, 33,* 14-27.

Brueggemann, A., & Gable, S. (2018). Preschoolers' selective sustained attention and numeracy skills and knowledge. *Journal of Experimental Psychology, 171,* 138-147.

Brynner, J. (2004). Literacy, Numeracy And Employability: Evidence from the British birth cohort studies. *Literacy and Numeracy Studies, 13*(1), 31-48.

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

Chernyak, N., Harris, P. L., Cordes, S. (2019). Explaining early moral hypocrisy: Numerical cognition promotes equal sharing behavior in preschool-aged children. *Developmental Science, 22*(1), e12659.

Chernyak, N., Sandham, B., Harris, P. L., & Cordes, S. (2016). Numerical cognition explains age-related changes in third-party fairness. *Developmental Psychology, 52*(10), 1555-1562.

Chu, F. W., vanMarle, K., & Geary, D. C. (2015). Early numerical foundations of young children's mathematical development. *Journal of Experimental Child Psychology, 132, 205-212.*

Chu, F. W., vanMarle, K., & Geary, D. C. (2016). Predicting Children's Reading and Mathematics Achievement from Early Quantitative Knowledge and Domain-General Cognitive Abilities. *Frontiers in Psychology.* https://doi.org/10.3389/fpsyg.2016.00775 .

Chu, F. W., vanMarle, K., Rouder, J., & Geary, D. C. (2018). Children's early understanding of number predicts their later problem-solving sophistication in addition. *Journal of Experimental Child Psychology, 169,* 73-92.

Crooks, N. M., & Alibali, M. W. (2014). Defining and measuring conceptual knowledge in mathematics. *Developmental Review, 34*, 344-377.

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

Diaz, M. G., & Harris, P. L. (1990). The influence of the imagination on reasoning by young children. *British Journal of Developmental Psychology, 8,* 305-318.

Donaldson, M. (1978). *Children's minds.* Harper Perennial.

Gallistel, C. R., & Gelman, R. (1992). Preverbal and verbal counting and computation. *Cognition, 44*(1-2), 43-74.

Geary, D. C., vanMarle, K., Chu, F. W., Rouder, J., Hoard, M. K., & Nugent, L. (2018). Early Conceptual Understanding of Cardinality Predicts Superior School-Entry Number-System Knowledge. *Psychological Science, 29*(2), 191-205.

Gelman, R. (1993). A Rational-Constructivist Account of Early Learning About Numbers and Objects. *Psychology of Learning and Motivation, 30,* 61-96.

Gelman, R., & Gallistel, C. R., (1978). *The child's understanding of number.* Harvard University Press.

Gelman, R., & Meck. E. (1983). Preschoolers' counting: Principles before skill. *Cognition, 13,* 343-359.

Halford, G. S., Andrews, G., & Bowden, D. (1998, July 1-4). *Relational Complexity and Theory-of-Mind* [Poster presentation]. Biennial Meetings of the International Society for the Study of Behavioural Development, Berne, Switzerland. https://eric.ed.gov/?id=ED421673 .

Harrer, M., Cuijpers, P., Furukawa, T.A, & Ebert, D. D. (2019). *Doing Meta-Analysis in R: A Hands-on Guide*. DOI: 10.5281/zenodo.2551803. https://bookdown.org/MathiasHarrer/Doing_Meta_Analysis_in_R/

Higgins, J. P. T., Thompson, S.G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *British Medical Journal, 327,* 557-560.

Higgins, J. (2016). *Explaining heterogeneity* [PowerPoint slides]. Methods.Cochrane.org. https://methods.cochrane.org/statistics/sites/methods.cochrane.org.statistics/files/public/uploads/SMG_training_course_2016/sess2.pdf .

Hyde, D. C., Simon, C. E., Berteletti, I., & Mou, Y. (2017). The relationship between non-verbal systems of number and counting development: a neural signatures approach. *Developmental Science, 20*(6), e12464.

Jordon, N. C., Kaplan, D., Ramineni, C., & Locuniak, M. N. (2009). Early Math Matters: Kindergarten Number Competence and Later Mathematics Outcomes. *Developmental Psychology, 45*(3), 850-867.

Koponen, T., Salmi, P., Eklund, K., & Aro, T. (2013). Counting and RAN: Predictors of arithmetic calculation and reading fluency. *Journal of Education Psychology, 105*(1), 162-175.

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*(2), 130-169.

Lee, M. D., & Sarnecka, B. W. (2010). A model of knower-level behaviour in number concept development. *Cognitive Science, 34,* 51–67.

Lee, M. D., & Sarnecka, B. W. (2011). Number-knower levels in young children: Insights from Bayesian modelling. *Cognition, 120,* 391-402.

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 Psychology, 195,* 104846

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & The PRISMA Group. (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLoS Med 6*(7): e1000097. https://doi.org/10.1371/journal.pmed.1000097

Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., Shekelle, P., Stewart, L. A., & PRISMA-P Group. (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Systematic Reviews, 4*(1). http://www.systematicreviewsjournal.com/content/4/1/1

Novianti, P. W., Roes, K. C. B., & van der Tweel, I. (2014). Estimation of between-trial variance in sequential meta-analyses: A simulation study. *Contemporary Clinical Trials, 37,* 129-138.

Nunes Carraher, T., Carraher, D. W., & Schliemann, A, D. (1985). Mathematics in the streets and schools. *British Journal of Developmental Psychology, 3,* 21-29.

Ouzzani, M., Hammady, H., Fedorowicz, Z., & Elmagarmid, A. (2016). Rayyan — a web and mobile app for systematic reviews. *Systematic Reviews, 5*(210), DOI: 10.1186/s13643-016-0384-4.

Page M. J., McKenzie J. E., Bossuyt P. M., Boutron I., Hoffmann T. C., Mulrow C. D., Shamseer, L., Tetzlaff, J. M., Akl, E.A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., . . . Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ,372*(71). doi:10.1136/bmj.n71

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.

Reyna, V. F., Nelson, W. L., Han, P. K., & Dieckmann, N. F. (2009). *Psychological Bulletin, 135(6), 943-973.*

Sarnecka, B. W., & Carey, S. (2008). How counting represents number: what children must learn and when they learn it. *Cognition, 108,* 662-674.

Sarnecka, B. W., & Lee, M. D. (2009). Levels of number knowledge during early childhood. *Journal of Experimental Child Psychology, 103*, 325-337.

Staaks, J. (2020, Oct 7). *Systematic Review Search Support* [PowerPoint Slides]. OSF. https://osf.io/49t8x/ .

Starr, A., Libertus, M. E., & Brannon, E. M. (2013). Number sense in infancy predicts mathematical abilities in childhood. *PNAS, 110*(45), 18116-18120.

Topor, M., Pickering, J. S., Barbosa Mendes, A., Bishop, D. V. M., Büttner, F. C., Henderson, E. L., … Westwood, S. J. (2020). Non-Interventional, Reproducible, and Open (NIRO) Systematic Review guidelines v1. https://doi.org/10.17605/OSF.IO/F3BRW .

Verdine, B. N., Golinkoff, R. M., Hirsh-Pasek, K., Newcombe, N. S., Filipowicz, A. T., & Chang, A. (2014). Deconstructing building blocks: Preschoolers' spatial assembly performance relates to early mathematical skills. *Child Development, 85*(3), 1062-1076.

Veroniki, A. A., Jackson, D., Viechtbauer, W., Bender, R., Bowden, J., Knapp, G., Kuss, O., Higgins, J. P. T., Langan., & Salanti, G. (2015). Methods to estimate the between-study variance and its uncertainty in meta-analysis. *Research Synthesis Methods, 7,* 55-79. DOI: 10.1002/jrsm.1164 .

Wagner, J. B., & Johnson, S. C. (2011). An Association between understanding cardinality and analog magnitude representations in pre-schoolers. *Cognition, 119*(1), 10 -22.

Wang, N. (2018). *How to conduct a Meta-Analysis of Proportions in R: A Comprehensive Tutorial.* [Unpublished manuscript]. Department of Educational Psychology, Texas A&M University. https://doi.org/10.13140/RG.2.2.27199.00161 .

Weisberg, D. S., & Gopnik, A. 2013. Pretense, counterfactuals and Bayesian causal models: Why what is not real really matters. *Cognitive Science, 37,* 1368-1381.

Wilkey, E. D., & Ansari, D. (2020). Challenging the neurobiological link between number sense and symbolic numerical abilities. *ANNALS of the New York Academy of Sciences, 1464*(1), 76-98.

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

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

Experimental Chapter 2: Examining the Influence of Pretend Play on Children's Displays of Cardinal Proficiency

The previous chapter presented a meta-analytic exploration of the Give-N task, as the primary measure of cardinal proficiency. The analysis suggested that this measure may not be robust and different implementations of the task may affect children's displays of cardinal proficiency. This highlighted the needs both for standardisation of this task and for current theories of cardinal proficiency development to acknowledge these effects.

Importantly, we must acknowledge these effects and begin to explore why these differences in implementation influence children's displays of cardinal proficiency. Not only will this make current theoretical explanations of cardinal proficiency more complete, but this may provide more detail as to *how* this important skill develops. The previous study suggested some implementation methods facilitated children's displays of cardinal proficiency, whilst others hindered their performance. For example, asking for numbers of items in a random order reduced children's pass rate. However, as the previous study was meta-analytic it is difficult to establish if there is systematicity in the findings or if these influences are a result of unknown variables not recorded within the methods of the research papers. These findings need to be explored experimentally, and this is the focus of the next chapter.

One key finding from the previous chapter suggested that when the Give-N task was presented in a role-play / pretend play format, this facilitated children's displays of cardinal proficiency. Our focus for the next chapter centred on this finding, as opposed to how the children were asked for the items, for example. We believed the pretend play element may provide a more in-depth insight into *how* the cardinal principle develops. However, as discussed above, this needed to be tested experimentally, to explicate why this might be.

Literature exploring pretend play across many areas of developmental psychology suggests that pretend play may free children from the rules of everyday life, allowing them to perform differently on everyday tasks (Diaz & Harris, 1990; Weisberg & Gopnik, 2013). Likewise, it may be that naturalistic situations, more representative of children's daily lives may facilitate performance (Nunes Carraher et al., 1985) or that presenting a task in a more meaningful way may enhance performance (Donaldson, 1978, 1982). These different explanations are explored experimentally in the next chapter, where we presented the Give-N task to children in three different ways reflecting the above explanations.

The next chapter contains the second paper in this journal format thesis, and explores the original question experimentally: *Is the Give -N task, the primary measure of cardinal proficiency, a reliable and robust measure?*

The Effects of Pretend Play on Children's

Displays of Cardinal Proficiency.

Hannah Thomas

Charlie Lewis

Kirsty Dunn

Contributions

HT, CL & KD designed the study, with HT and CL writing the manuscript. HT recruited and children for the experiment, carried out the experiments, handled and coded the data, and performed statistical analyses.

Abstract

Development of the cardinal principle is a key milestone in children's understanding of number, here children understand the final number in a count represents the total number of items. Recent research suggests that different implementations of the primary measure of cardinal proficiency may affect children's performance on this task. Within this study we test this experimentally. Adopting a within-subject design we presented children with the standard task, a role-play version of the task and a naturalistic version of the task. Although the results did not reach significance the findings suggest that different implementations of the same task, do affect children's displays of cardinal proficiency, just not in a systematic way. The findings are discussed in relation to current theoretical positions and the importance of these effects when assessing children's cardinal proficiency.

Introduction

Cardinal proficiency, understanding that the final number in a count represents the total number of items, is a key competency for young children and is an important foundation for future mathematical achievement (Chu et al., 2016; Geary et al., 2018). Typically, development of the cardinal principle is measured using the Give-N task (Crooks & Alibali, 2014). However, recent research suggests that presenting this task in different ways may influence children's displays of cardinal proficiency (Thomas et al., to be submitted: chapter one). This raises questions regarding the reliability and the validity of the Give-N task, and has significant implications for theories of cardinal proficiency development that rely upon this task as their primary measure. Here we test the findings from the meta-analysis presented in chapter one and explore these experimentally, examining the effects of administering this task using three different approaches.

Typically, the Give-N task requires children to give or produce a sub-set of items from a larger set. Children are deemed to have passed the task, and are credited with cardinal proficiency when they are able to give five or six (or more) items. Use of the Give-N task, as a measure of cardinal proficiency and mathematical competency, is ubiquitous throughout the developmental literature. The task permeates nearly every field of developmental psychology. From exploring relationships between mathematical development and executive functions, language abilities (Purpura & Ganley, 2014), spatial skills (Verdine et al., 2014), sharing capabilities (Chernyak et al., 2016; Chernyak et al., 2019) and selective attention (Brueggemann & Gable, 2018), to how home numeracy activities (Soto-Calvo et al., 2020) or spontaneous focus on number (Hannula et al., 2007) can improve mathematical proficiency. The extensive use of the Give-N task, throughout many research fields raises one question, is this task a reliable and valid measure of cardinal proficiency? Indeed, this crucial question, and lack of an empirical answer has been highlighted recently (Sella et al., 2021).

A systematic review and meta-analysis by Thomas et al., (to be submitted: see previous chapter) begins to address this question. This highlights the numerous ways the Give-N task has been presented to children, and the many different ways in which they are credited with cardinal proficiency during this task. Importantly, the review suggests that presenting the Give-N task during role-play activities may facilitate children's displays of cardinal proficiency. This finding is significant. If children's displays of cardinal proficiency can be influenced in a systematic way this limits the validity of the Give-N task, and brings into question the theories and findings based on the task. The meta-analysis has substantiated the concerns of Sella et al., (2021), where questions were posited regarding the reliability of the Give-N task and that different versions may produce different results. However, the meta-analysis amalgamates existing evidence, with no regulation over the studies' exact procedures, therefore this finding needs to be tested experimentally.

Furthermore, this has direct applicability to current theories of cardinal proficiency development. Within the literature there are two opposing views of cardinal proficiency development, the principles-first view and the principles-after. In brief, advocates of a principles-first view take a continuity stance, believing that innate counting principles underpin this important skill and cardinal proficiency is reached through superior counting procedures (Gallistel & Gelman, 1992; Gelman 1993; Gelman & Meck, 1983). Whereas, those who subscribe to a principles-after or discontinuity stance consider cardinal principle acquisition to represent a conceptual change (Carey, 2004; Lee & Sarnecka, 2010). Both theoretical stances focus on the developmental pattern of cardinal principle acquisition, and both utilise the Give-N task as their primary measure of cardinal proficiency. However, neither theoretical account considers how the presentation of this task may influence children's displays of cardinal proficiency. If, as the meta-analysis (Thomas et al., to be submitted: see previous chapter) suggests, presenting the Give-N task during role-play activities does influence children's displays of cardinal proficiency then this brings into question the patterns of development suggested by both theoretical positions.

Moreover, if the presentation of this task affects children's displays of cardinal proficiency this poses a significant problem for the principles-after view of cardinal principle acquisition, specifically the Knower-Levels account. A key tenet within the principles after account of cardinal proficiency development is that cardinal principle acquisition represents a conceptual change, and therefore once children have grasped this principle their performance across tasks should be equivalent (Lee & Sarnecka, 2010, 2011; Le Corre et al., 2006; Sarnecka & Lee, 2009). If, when tested experimentally, different presentations of the same task have an effect on children's performance, then this may raise questions regarding their premise that cardinal principle development is a sudden conceptual change. Likewise, if children's performance on other equivalent cardinality tasks, this again raises questions regarding the knower-levels account of development.

There are equivalences between the questions raised here and the criticisms made against traditional Piagetian number conservation tasks (Donaldson, 1982). During these tasks children are asked a numerical equivalence question, before and after a transformation. Typically, children under the age of six give an incorrect response after the transformation. However, when these tasks are presented in a way which is meaningful to the child, more children produce correct responses (Light et al., 1979; McGarrigle & Donaldson, 1975). It is suggested that presenting this task in an abstract way, with no specific purpose or reason for the transformation may confuse or mislead the child. They may be unable to comprehend the motives of the researcher and the demands of the task (Donaldson, 1978). This would reduce the validity of the procedure, and it may be that such tasks are not an appropriate measure for this specific skill (Donaldson, 1978). The possibility this may also apply to the Give-N task must be considered.

The traditional Give-N task is presented with limited purpose, and the abstract nature of the task may leave children wondering what the researcher requires of them, limiting their displays of cardinal proficiency. Theoretically, the change we see as children begin to pass the Give-N task may not be a conceptual one, merely that with age or practice children are able to manage a more abstract situation. This does not mean that cardinal principle acquisition is not a conceptual change, just that the Give-N task, as with traditional conservation tasks, may not be an appropriate tool for measuring cardinal principle development, and this directly impacts upon theories of cardinal proficiency development when using this task. It is therefore important we test the findings from the meta-analysis (Thomas et al., to be submitted: see previous chapter) experimentally, and clarify if role-play activities do influence children's displays of cardinal proficiency.

It is also important we consider the reasons why role-play activities may facilitate children's displays of cardinal proficiency. It is possible that naturalistic situations, typical of children's daily lives, may recruit different routines that facilitate cardinal proficiency (Nunes Carraher et al., 1985). Alternatively, it may be that engaging in pretend play activities may free children from the 'rules' of reality allowing them to reason about number in different ways (Diaz & Harris, 1990; Weisberg & Gopnik, 2013). Or, as discussed, framing the traditional Give-N task in ways that are meaningful to the child may make the task more comprehendible, thus improving children's displays of cardinal proficiency (Donaldson, 1978, 1982).

Evidently the validity of the Give-N task is uncertain. Equally, the theoretical positions that adopt this task as their primary measure are insufficient, as they do not account for how changes in task presentation may affect children's displays of cardinal proficiency. It is therefore important to re-examine the theoretical positions, and to explicate the possible reasons why pretend play may facilitate children's displays of cardinal proficiency. In response to this, and the preceding meta-analysis, we developed a within-participant experimental design whereby children's cardinal proficiency is examined during three different versions of the Give-N task, along with the What's on the Card task as a cardinal proficiency equivalent task. The three versions of the Give-N task include a role-play version, a naturalistic version, and the standard traditional Give-N task. Three alternative hypotheses are tested. Based on evidence suggesting that role-play can free children from the 'rules' of reality (Diaz & Harris, 1990; Weisberg & Gopnik, 2013), it is hypothesised that children's displays of cardinal proficiency will be significantly higher during the role-play version of the Give-N task than during the naturalistic or traditional version of this task. In contrast, based on the suggestion that naturalistic situations, typical of children's daily lives, may recruit different routines that facilitate cardinal proficiency (Nunes Carraher et al., 1985), it is hypothesised that children's displays of cardinal proficiency will be significantly higher during the naturalistic version of the Give-N task than during the role-play or traditional version of this task. Alternatively, based on the evidence that framing tasks in a meaningful way can facilitate children's achievements (Donaldson, 1978, 1982) it is hypothesised that children's displays of cardinal proficiency will be significantly higher during both the naturalistic version and the role-play version of the Give-N task, than during the traditional Give-N task.

Method

Pre-registration

The study was pre-registered with Open Science Framework (https://osf.io/bp5ec/) on the 23 July 2021.

Participants

In line with other developmental studies and reflecting on the current testing availabilities following Covid-19, our aim was to recruit 65 children. This being the equivalent of 20 participants per group (should this have been a between subjects design) allowing for some expected exclusions or drop out. Children were recruited from four private local nurseries, all within the Morecambe district. In total 66 children were recruited, however, 14 did not complete all three testing sessions and were therefore excluded. The remaining 52 children were between 27 and 55 months M_{age} = 44.23 months, SD_{age} = 6.93. The mean age of children across the four different nursery settings was comparable (see Table S1).

Materials

Traditional Give-N task

For this task a medium sized cuddly toy rabbit, named Rachel, and a plastic tub were used for children to place different animals on and 'give' these to Rachel. The animals comprised three sets of small plastic animals; 15 frogs, 15 pigs and 15 ducks. Each set of animals was contained in a separate plastic tray / basket.

Naturalistic Give-N task

A length of string was utilised as a washing line by the researcher to hang up different items of dollies clothing. These consisted of small dresses, trousers, t-shirts, cloth nappies and towels, and were contained in a small washing basket. Children were given 15 plastic dolly pegs, in a basket, to help the researcher hang up the washing.

Role Play Give-N task

For this task a small table was utilised as a 'market stall'. The researcher acted as a customer and held a small plastic basket into which children were asked to place different fruits. The fruits consisted of 15 small plastic apples, 15 small plastic oranges and 15 small plastic bananas. Each set of fruit was contained in an individual tray / basket, and these were situated on the 'market stall'. The researcher also had 18 small plastic coins (contained in a drawstring pouch) to pay for the fruits, and the children had a plastic pot to put these in.

What's on the Card task

A set of A4 laminated cardboard cards were used for this task. Printed on the cards were different sets of animals. For example, one card displayed five pigs (see Figure S1). In total there were 18 cards displaying different sets of animals plus one demonstration card (see Table S2 for a full list of the different cards)

Procedure

The three Give-N tasks were presented to children following a Latin Square design. The What's on the Card task was completed last. Each task was completed on separate days, and all the tasks took place in a quiet room at the child's nursery or preschool. All the tasks were video-recorded for offline coding, with the number of items given by the child recorded for the Give-N trials, and the number word used by the child recorded for the What's on the Card trials.

For all the Give-N tasks the communication between the researcher and the child was kept as similar as possible. In addition, the Give-N tasks followed an almost identical procedure to maintain parity across the tasks. See supplementary materials for the script the researcher followed for the tasks.

Traditional Give-N task

Children were seated at a table, with the researcher opposite. On the table in front of the researcher was Rachel the rabbit and the plastic plate. The three baskets of animals were placed next to the child on the table. The children were introduced to Rachel the rabbit and were told "today you're going to give her animals to play with". After this, children received one practice trial, where they were asked to give Rachel one pig and to pop it on her plate. Following the practice trial children received feedback; either "Ooh lovely, thank you" if they gave the correct number of pigs or, if they gave the incorrect number of pigs "Oh I don't think that's just one pig, Rachel only wants one pig. Can you give her just one pig please?"

If children failed the practice trial for a second time the researcher moved onto the test trials irrespectively. For those who succeed on the practice trial either first or second time the test trials immediately followed their success.

During the test trials children were asked to give Rachel the rabbit varying numbers of either pigs, cows, or horses. The numbers requested were 2, 3 and 5 in the first block and 6, 8 and 10 in a second block. Each number was asked for three times, and the order in which the numbers were requested was randomised within their blocks, and was different for each child. In addition, the type of animal requested on each trial was randomised (see supplementary materials for a scoring sheet illustrating the randomisation procedure). The animals were returned to their specific baskets after each trial to maintain the same set size from which children selected the animals. No feedback was given to children during test trials, the researcher responded only with "Ooh lovely, thank you" after the child had passed the animals, regardless of the number of animals given.

Naturalistic Give-N task

For this task the child stood with the researcher. The washing line was hung in front of the researcher, the basket of dollies clothes next to the researcher and the basket of pegs on floor near the child. The task began with the researcher asking if the child could help her hang up some washing. For the practice trial the researcher picked up an item of dollies clothing and asked the child "Can I have one peg please", and held out their hand. Following the practice trial feedback was provided and test trials commenced as detailed for the Traditional Give-N task.

During the test trials the researcher picked up a piece of washing and asked for varying numbers of pegs. The blocks of numbers, and the order in which they were requested mirrored those in the Traditional Give-N task. Again, the feedback remained constant "Ooh lovely, thank you" regardless of their response, and more pegs were discretely added to the child's basket after each trial to preserve the set size of 15 from which the child selected pegs.

Role Play Give-N task

The child stood with researcher, while the researcher introduced the shopping game. The 'market stall' with the baskets of fruit on was situated near the child. The plastic pot for children to deposit the coins was also placed on the 'market stall'. The researcher stood near the child holding the plastic shopping basket, and their supply of coins.

The task began with a practice trial, where the researcher greeted the child as the 'shopkeeper' and asked for one apple to be put in their shopping basket. Following the practice trial feedback was provided and test trials commenced as detailed for the Traditional Give-N task.

During the test trials the researcher asked for various numbers of different fruits, and handed the child a coin in payment following their response. The blocks of numbers, and the order in which they were requested mirrored those in the Traditional Give-N task, and the type of fruit requested on each trial was randomised. Again, the feedback remained the same as the other Give-N tasks, and after each test trial the fruits were returned to their respective baskets to maintain the appropriate set sizes.

What's on the Card task

For this task children were seated at a table. Following a similar procedure to Sarnecka & Negen (2019, https://osf.io/eznht/) the task began with the researcher showing the child the demonstration card, and saying "look here are some cards with pictures of cows, pigs and horses on". The practice trial followed with the researcher showing the child a card with one pig on and asking "What's on this card?". If the child replied with "A pig" the researcher explained that in this game we use number words, so we say "One pig", the child was asked to repeat this. The test trials followed with the researcher showing the child all 18 test cards individually, and asking "What's on this card?". The order the cards where shown to the child mirrored the block and number randomisation in the Traditional Give-N task, which was individual for each child.

Coding

Offline video coding was completed by the primary researcher, with 20% of responses data second coded by a trained research assistant with 100% inter-rater reliability.

Results

Analysis

All analyses were conducted using R and R studio version 4.3.3.

Knower-Levels and Cardinal Principle Knowers

Within the sample of 52 children there were varying degrees of cardinal proficiency across the tasks. See Table 1 for children's knower-levels across the four cardinal proficiency tasks, and Table 2 for numbers of children, across the different tasks, who were able to give five or more (and therefore classified as cardinal principle knowers). We also examined the number of cardinal principle knowers across the different nursery settings (see Table 3).

Table 1

Number of children achieving each knower-level split by task

	Knower-level achieved						
Task	0	2	3	5	6	8	10
Traditional	5	15	16	2	3	3	7
Role Play	6	12	18	2	2	3	8

Naturalistic	4	14	15	3	6	2	7
What's on the Card	6	5	10	12	2	7	9

Table 2

Number of children achieving cardinal principle status split by task

	Cardinal principle status achieved	
Task	Yes	No
Traditional	15	37
Role Play	15	37
Naturalistic	18	34
What's on the Card	30	22

Table 3

Number of cardinal principle knowers across the different nursery settings, split by task

	Number of cardinal principle knowers							
	Traditional Task		Role Play Task		Naturalistic Task		What's on the card	
Nursery Setting	Yes	No	Yes	No	Yes	No	Yes	No
A	3	5	3	5	3	5	4	4
B	6	8	6	8	6	8	10	4
C	2	8	3	7	2	8	6	4
D	4	16	3	17	3	17	8	12

Assessment of Knower-Levels

To assess if different presentations of the Give-N task influenced children's displays of cardinal proficiency an ordinal cumulative link model was undertaken using the Ordinal package in R. Adopting a maximal approach a CLM was built with the effects of task presentation (including the What's on the Card task), order of presentation, age of children, and nursery setting entered as predictor variables for children's knower-levels (see Table 4).

There were three significant predictors of children's knower-level knowledge. These were age in months, nursery setting, and task type (see Table 4 for model output). As expected, older children displayed higher knower-level knowledge. However, an effect of nursery setting suggested that within nursery setting D children's displays of knower-level knowledge were significantly lower.

To confirm the difference in performance across tasks (naturalistic, role play, traditional & WOC), and to correct for multiple comparisons post-hoc analyses were undertaken for task presentation / type. These revealed no significant differences between any of the tasks. Subsequently, Spearman's rank order correlations were undertaken to assess the relationship between each task. This revealed significant correlations between all of the tasks see Table 5.

Table 4

Ordinal regression model showing the effects of age, order of presentation, nursery setting, and task

	B	*SE*	*Z*
Role Play	-.16	.36	-.26
Traditional	-.20	.36	-057
What's on the Card	.74	.36	2.05*
Latin order 2	.58	.32	1.83
Latin order 3	.34	.31	1.10
Nursery D	-1.63	.36	-4.50***

Nursery A	-.46	.45	-1.03
Nursery B	.39	.38	1.02
Age	.21	.02	9.17***
	Model Fit		
	Nagelkerke R^2	LL	χ^2
	.45	-61.68	123.36***

R model equation: knower-levels ~task + latin + nursery + age

*p<.05 ***p<.001

Table 5

Correlation matrix showing relationships between knower-levels across the different task versions

Task				
	Traditional	Naturalistic	Role Play	WOC
Traditional	/			
Naturalistic	.72*	/		
Role Play	.76*	.86*	/	
WOC	.74*	.69*	.67*	/

* *p < .001*

Assessment of cardinal principle knowledge

To assess if different presentations of the Give-N task influenced children's cardinal principle knowledge (i.e., those who are able to give five items or more) a binomial logistic regression model was fitted using the LME4 package in R. Adopting a maximal approach a GLM was built with the effects of task presentation (including What's on the Card task), order of presentation, age of children, and nursery setting entered as predictor variables for children's knower-levels.

Again, age in months, nursery setting, and task presentation were significant predictors of children's cardinal principle knowledge. In addition, order of presentation was significant within this model (see Table 6 for model output). As before, older children displayed higher cardinal principle knowledge. However, when exploring task presentation and cardinal principle knowledge (opposed to knower-levels), post hoc Bonferroni-Holm analyses revealed higher displays of cardinal proficiency during the what's on the card task than were seen during the role play or traditional task (see Table 7). Likewise, nursery setting D had fewer children with cardinal principle knowledge. However, when examining cardinal principle knowledge specifically, this suggests that children presented with either the Latin square order two (naturalistic task, role play task then the traditional task) or three (role play task, traditional task then naturalistic task) displayed higher levels of cardinal proficiency. Furthermore, nursery setting B had more children with cardinal principle knowledge. Again, there was no interaction between task and nursery setting.

Table 6

Binary logistic regression model showing the effects of age, order of presentation, nursery setting, and task on cardinal principle knowledge

	B	*SE*	*Z*
Intercept	-11.34	1.94	-5.84
Role Play	-.39	.51	-.76
Traditional	-.39	.51	-.76
What's on the Card	1.15	.49	2.33*
Latin order 2	1.05	.50	2.10*
Latin order 3	1.03	.44	2.34*
Nursery D	-1.10	.50	-2.20*
Nursery A	.50	.59	.86
Nursery B	1.49	.53	2.81***
Age	.22	.04	5.60***
		Model Fit	
	Nagelkerke R^2	LL	χ^2
	.44	-40.12	80.25***

R model equation: cardinal principle knowledge ~task + latin + nursery + age

*p<.05 ***p<.001

Table 7

Post hoc Bonferroni-Holm comparisons for task presentation

Task	B	*SE*	*Z*	*P*
Naturalistic / Role Play	.39	.51	.76	1.00
Naturalistic / Traditional	.39	.51	.76	1.00
Naturalistic / WOC	-1.15	.49	-2.33	.08
Role Play / Traditional	.00	.52	.00	1.00
Role Play / WOC	-1.53	.51	-3.00	.02
Traditional / WOC	-1.53	.51	-3.00	.02

First task analysis

As the Latin square order was a significant predictor variable in the above model, we examined the first and last task data individually. This was modelled as above with age, order, nursery setting as predictor variables. However, there was no longer an effect of order (see supplementary materials for output).

Discussion

As expected, age was a significant predictor of children's cardinal proficiency displays. With older children showing greater levels of cardinal principle knowledge. None of the other hypothesised outcomes for this study were supported. We found no evidence to suggest that role play, or naturalistic contexts facilitate children's displays of cardinal proficiency. However, subsidiary findings do suggest that children's displays of cardinal proficiency are greater during a comparable task (What's on the Card). Likewise, some nursery settings appear to have higher levels of cardinal proficiency, despite there being no difference in the ages of children attending each setting.

Role Play vs Natural Contexts

Whilst we did see some within participant differences in the raw data between performance on the traditional Give-N task and the alternative versions this did not reach significance. This was unexpected following the results of the meta-analysis (Thomas et al., to be submitted: see previous chapter). Likewise, with reference to the literature surrounding natural contexts and engagement (Diaz & Harris, 1990; Light et al., 1979; McGarrigle & Donaldson, 1975; Weisberg & Gopnik, 2013) we expected to see enhanced displays of cardinal proficiency during the alternative tasks.

It is possible that the effect of role play seen in the meta-analysis could be a spurious finding, as many of the studies adopting this methodology were from the same research group (Sella et al., 2017; Sella et al., 2018; Sella et al., 2019; Sella & Lucangeli, 2020; Sella et al., 2020). Therefore, the effect could have been a researcher-based engagement effect, with an individual researcher being particularly engaging. However, there were a number of role-play studies in the meta-analysis and it is unlikely all were undertaken by the same researcher. Nevertheless, this could explain our results here. All the testing sessions were undertaken by the primary researcher. Maybe her style was particularly engaging (or distracting) and therefore reduced the effects of role-play or natural contexts. Or perhaps engagement is the sole effect and the two, potentially more engaging, versions of the task diluted the results? Should we have only had one alternative Give-N task our results may have reached significance. However, additional research is needed to explicate this.

Likewise, our results may be a due to a lack of power, having only 52 participants in the study (equivalent to approximately 17 per group, should the study have been between-participants design). It is possible that the trends seen within the data may have reached significance should we have had a larger participant group.

Whilst the results from this study were not significant these, coupled with the meta-analysis in chapter one, may still contribute to continuing discussions. Engagement may still be a factor to consider when assessing children's cardinal principle knowledge, and this is something the primary accounts of cardinal proficiency development do not acknowledge. Further research may help to elucidate why engagement, or role-play, or natural contexts may have an effect on children's displays of cardinal proficiency, and therefore enhance the current description of cardinal proficiency development. A deeper understanding of these effects would help shift current theory from a descriptive level to an explanation of *how* this vital skill develops.

What's on the Card Task

This subsidiary finding is important for three reasons. First, the primary knower-levels account of cardinal proficiency asserts, as a central tenet, that when children have acquired the cardinal principle (i.e., when children are able to give five items), this represents a conceptual change and therefore this knowledge is generalised to all other cardinal proficiency tasks (Lee & Sarnecka, 2010, 2011; Le Corre et al., 2006; Sarnecka & Lee, 2009). This is not supported by our results. Children's displays of cardinal proficiency were significantly greater on this task, than any version of the Give-N task. When we examine the cardinal principle knowers, 30 children were classified as knowers on this task, whereas only 18 were classified as cardinal principle knowers on the naturalistic task, and only 15 on the role-play and traditional task attained this level of performance. This casts doubt on the assertion that cardinal principle development is a sudden conceptual change when children can give five items. The knower-levels account of cardinal proficiency development may need revision.

Whilst this result does not necessarily correspond directly with emerging evidence suggesting that cardinal principle development is a gradual change (Krajcsi & Reynvoet, 2023), it suggests that a gradual development of the cardinal principle is more likely as children's displays of this skill are not consistent across tasks.

Secondly, this result may correspond with our original hypotheses regarding natural contexts. When we examine the knower-levels data, we also find that knower-levels are significantly higher on this task than any of the other Give-N tasks. Perhaps this more 'paper-based' task reflects children's experiences, give that tasks like this are close to our current curriculum for early years approaches to teaching early mathematical concepts. Often, counting skills are capitalised upon during reading books, looking at numerals on the wall, or counting concrete items on a table, as opposed to asking children to give numbers of items. The Early Years Foundation Stage curriculum suggests that counting experiences for practitioners should be incorporate within their daily routines centred around counting objects or fingers and toes, rather than giving items (Department for Education, 2024). Perhaps, the enhanced performance on the What's on the Card task within this study is reflective of children's daily routines and therefore an effect of natural contexts as suggested by (Nunes Carraher et al., 1985). If so, this is something that current accounts of cardinal proficiency development do not consider, and this would again suggest that the knower-levels account of cardinal proficiency development needs revision.

Finally, the improved performance on this alternative measure of cardinal proficiency highlights limitations within this field of research. The over-reliance on the Give-N task as the primary measure of cardinal proficiency is problematic, as the results here suggest that children's displays of cardinal proficiency may vary from task to task. Whilst this could be due to the procedural demands of different tasks, it could also reflect current curriculum with nursery settings. Either way it is something that we must be aware of when measuring this foundational skill. Perhaps, triangulation is applicable here and we need to assess children's skills in a multi-faceted way.

Nursery Effect

Another subsidiary finding was that some nursery settings had far fewer cardinal principle knowers than others and children's displays of cardinal proficiency were also significantly lower (i.e., the knower-levels classifications). There could be many reasons for this effect. It could be reflective of the differences in socio-economic status (SES) in the catchment areas for each nursery. However, many of these nurseries are within close locality, so this may not apply or be difficult to differentiate. Likewise, the effect of nursery could reflect subtle differences in the early number skills provision within each setting. However, without further examination and observations any statement to this effect would be anecdotal. Equally, further examination could be difficult, due to the inter-play between nursery provision, numbers of special educational needs children at each setting, the exact SES for each setting, and also a knowledge of the home maths experience for the children. This is a complex set of dynamics to tease apart.

Task Order Effect

The significant effect of task order suggested that children who received the traditional task first displayed lower overall cardinal principle knowledge than those who received either the naturalistic task or the role play task first. This suggests that there may be a subtle interaction between the type of task and the order in which they are presented. It may be that, despite the gap in days between the testing sessions, some carry over effects remain. As discussed, we did see slight differences in children's performance across the raw data, with some children's performance superior on the traditional task, some on the naturalistic task and some on the role-play task. Should we have only had one alternative Give-N task our effects of engagement may have been significant. In essence, the engagement factor may have been diluted by having two alternative, potentially more engaging tasks. Therefore, perhaps the children who received the more engaging tasks first found any failures easier to recover from, resulting in improved performance on the subsequent Give-N variations. Likewise, this may be why our follow up first trial analysis did not produce any significant results.

As all children's displays of cardinal proficiency were superior on the What's on the Card task (and this was completed last), it is worth noting that the carry over effects may not be present here as this task is significantly different to the Give-N (see above for discussion).

Conclusion

The mixed results from this study may not have supported our original hypotheses. However, children's engagement during cardinal principle tasks is something for theory, practice and research to take forward. Whilst the within-subject, Latin square design has improved the validity of the study, our results may be limited due to the small area for participant recruitment, particularly when considering the possible curriculum effect and the What's on the Card finding. Nevertheless, this is still something current accounts of cardinal proficiency do not consider.

Overall, the results from this study suggest that sole use of the Give-N task as a measure of cardinal proficiency may not be reliable. Likewise, they raise questions for a central tenet of the primary account of cardinal proficiency development, and this needs further examination. Corresponding with emerging evidence in support of a gradual development of the cardinal principle (Krajcsi & Reynvoet, 2023), we need to explore *how* this foundational mathematical skill develops and enhance the current description of cardinal principle development.

References

Brueggemann, A., & Gable, S. (2018). Preschoolers' selective sustained attention and numeracy skills and knowledge. *Journal of Experimental Psychology, 171,* 138-147.

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

Chernyak, N., Harris, P. L., Cordes, S. (2019). Explaining early moral hypocrisy: Numerical cognition promotes equal sharing behavior in preschool-aged children. *Developmental Science, 22*(1), e12659.

Chernyak, N., Sandham, B., Harris, P. L., & Cordes, S. (2016). Numerical cognition explains age-related changes in third-party fairness. *Developmental Psychology, 52*(10), 1555-1562.

Chu, F. W., vanMarle, K., & Geary, D. C. (2016). Predicting Children's Reading and Mathematics Achievement from Early Quantitative Knowledge and Domain-General Cognitive Abilities. *Frontiers in Psychology.* https://doi.org/10.3389/fpsyg.2016.00775 .

Crooks, N. M., & Alibali, M. W. (2014). Defining and measuring conceptual knowledge in mathematics. *Developmental Review, 34*, 344-377.

Department for Education. (2024). Early years foundation stage profile handbook. Early years foundation stage profile handbook.

Diaz, M. G., & Harris, P. L. (1990). The influence of the imagination on reasoning by young children. *British Journal of Developmental Psychology, 8,* 305-318.

Donaldson, M. (1978). *Children's minds.* Harper Perennial.

Donaldson, M. (1982). Conservation: What is the question? *British Journal of Psycholgy, 73*(2), 199-207.

Gallistel, C. R., & Helman, R. (1992). Preverbal and verbal counting and computation.

Cognition, 44(1-2), 43-74.

Geary, D. C., vanMarle, K., Chu, F. W., Rouder, J., Hoard, M. K., & Nugent, L. (2018). Early Conceptual Understanding of Cardinality Predicts Superior School-Entry Number-System Knowledge. *Psychological Science, 29*(2), 191-205.

Gelman, R. (1993). A rational-constructivist account of early learning about numbers and objects. In D. L. Medin (ed.), *The Psychology of learning and motivation,* (pp. 61-96). Academic Press.

Gelman, R., & Meck. E. (1983). Preschoolers' counting: Principles before skill. *Cognition, 13,* 343-359.

Hannula-Sormunen, M. M., Rasanen, P., & Lehtinen, E. (2007). Development of counting skills: Role of spontaneous focusing on numerosity and subitizing-based enumeration. *Mathematical Thinking and Learning, 9*(1), 51-57.

Krajcsi, A., & Reynvoet, B. (2023). Miscategorised subset-knowers: Five and six knowers can compare only the number they know. *Developmental Science, 27*(1), e13430.

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*(2), 130-169.

Lee, M. D., & Sarnecka, B. W. (2010). A model of knower-level behaviour in number concept development. *Cognitive Science, 34,* 51–67.

Lee, M. D., & Sarnecka, B. W. (2011). Number-knower levels in young children: Insights from Bayesian modelling. *Cognition, 120,* 391-402.

Light, P. H., Buckingham, N., & Robbins, A. H. (1979). The conservation task as an interactional setting. *British Journal of Educational Psychology, 49*(3), 304-310.

McGarrigle, J., & Donaldson, M. (1975). Conservation accidents. *Cognition, 3,* 341-350.

Nunes Carraher, T., Carraher, D. W., & Schliemann, A, D. (1985). Mathematics in the streets and schools. *British Journal of Developmental Psychology, 3,* 21-29.

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.

Sarnecka, B. W., & Lee, M. D. (2009). Levels of number knowledge during early childhood. *Journal of Experimental Child Psychology, 103*, 325-337.

Sarnecka, B. W., & Negen, J. (2019). Longitudinal number-knower data. *Open Science Framework*, DOI 10.17605/OSF.IO/EZNHT

Sella, F., Slusser, E., Odic, D., Krajcsi, A. (2021). The emergence of children's natural number concepts: Current theoretical challenges. *Child Development Perspectives, 00,* 1-9.

Sella, F., Berteletti, I., Lucangeli, D., & Zorzi, M. (2017) Preschool children use space, rather than counting, to infer the numerical magnitude of digits: Evidence for a spatial mapping principle. *Cognition, 158,* 56-67.

Sella, F., Lucangeli, D., & Zorzi, M. (2018). Spatial and verbal routes to number comparison in young children. *Frontiers in Psychology, 9*(776).

Sella, F & Lucangeli, D. (2020). The knowledge of the preceding number reveals a mature understanding of the number sequence. *Cognition, 194*. 104104.

Sella, F., Lucangeli, D., & Zorzi, M. (2019). Spatial order relates to the exact numerical magnitude of digits in young children. *Journal of Experimental Child Psychology, 178*, 385-404.

Sella, F., Lucangeli, D., Kadosh, R. C., & Zorzi, M. (2020). Making sense of number words and arabic digits: Does order count more? *Child Development, 91*(5), 1456-1470.

Soto-Calvo, E., Simmons, F. R., Adams, A-M., Francis, H. N., Patel, H., Giofre. (2020)

Identifying the preschool home learning experiences that predict early number skills: Evidence from a longitudinal study. *Early Childhood Research Quarterly, 53,* 314-328.

Thomas, H. L., Lewis, C. & Dunn, K. (in preparation). A Systematic Review and Meta-analysis of the Give-N task as a measure of Cardinal Proficiency.

Verdine, B. N., Golinkoff, R. M., Hirsh-Pasek, K., Newcombe, N. S., Filipowicz, A. T., & Chang, A. (2014). Deconstructing building blocks: Preschoolers' spatial assembly performance relates to early mathematical skills. *Child Development, 85*(3), 1062-1076.

Weisberg, D. S., & Gopnik, A. 2013. Pretense, counterfactuals and Bayesian causal models: Why what is not real really matters. *Cognitive Science, 37,* 1368-1381.

Supplementary materials

Researcher script used for testing sessions

Practice trial Role Play:

"Let's play a shopping game . . . Look you are the shopkeeper and you have some carrots, tomatoes and potatoes to sell" (point to each corresponding vegetables, whilst saying this).

"I'll be the customer, who wants to buy some of your delicious vegetables! . . . . Good morning shopkeeper, please may have one carrot? Can you put it in my shopping basket please" (hold out the basket).

Feedback - if they give one carrot say "ooh lovely, thank you shopkeeper, here's the money" (hand over some coins).

if they don't give one carrot say "oh shopkeeper, I don't think that's just one carrot, I only want one carrot. Can you give me just one carrot please?"

TEST trials Role Play:

"Okay, thank you shopkeeper. Now, I think I'd like some X (veggies shown on the sheet) please may I have X (number of veggies shown on the sheet, and the type of veg shown on the sheet e.g., 3 potatoes) - (hold out the shopping basket)

"Ooh lovely, thank you shopkeeper, here's the money" (hand over coins).

Remember to replace the fruits i.e., put them back into the baskets after each trial.

Practice trial Traditional Give-N:

"Look, this is Rachel the rabbit, today you're going to give her animals to play with. You've got some ducks, pigs and frogs" (point to each corresponding animal, whilst saying this).

"Can you give Rachel one pig please? Pop it on her plate" (hold out the plate).

Feedback - if they give one pig say "ooh lovely, thank you".

if they don't give one pig say "oh I don't think that's just one pig, Rachel only wants one pig. Can you give her just one pig please?"

TEST trials Traditional Give-N:

"Okay, thank you. Now, Rachel would like some X (animals shown on the sheet) please may she have X (number of animals shown on the sheet, and the type of animal shown on the sheet e.g., 3 pigs) - (hold out the plate)

"Ooh lovely, thank you".

Remember to replace the animals i.e., put them back into the baskets after each trial.

Practice trial Naturalistic Give-N:

"Can you help me today? I've got some washing to hang up, and I need you to help pass me the pegs" (point to the pegs whilst saying this).

(Take hold of a piece of washing) "Can I have one peg please?" (hold out hand).

Feedback - if they give one peg say "ooh lovely, thank you".

if they don't give one peg say "oh I don't think that's just one peg, I only need one peg. Can you pass me just one peg please?"

TEST trials Naturalistic Give-N:

"Okay, thank you. This washing is dry I think (feel washing, take it down and put pegs back into child's basket). (Take hold of next piece of washing) now, I think I'll need some more pegs, please may I have X pegs (number of pegs shown on the sheet e.g., 3 pegs) - (hold out hand)

"Ooh lovely, thank you".

Remember to replace the pegs i.e., put them back into the basket after each trial.

Practice trial WOC:

"Look here are some cards with pictures of cows, pigs and horses on" (show cover card and point to each animal).

"So, what's on this card?" (show card with one pig on).

"That's right, but in this game we use our number words too. So, for this card you say ONE pig, can you say that?"

TEST trials WOC:

"Great, okay, so what would you say for this card?" (Show next card).

Example scoring sheet used during testing sessions

Latin Square	Date of Birth	Male / Female	Child ID
A, B, C			1

Count sequence

Highest number counted	
Additional notes about count sequence	

Give-N Traditional

Date:

Requested	Animal	Given
5	Pigs	
5	Frogs	
2	Pigs	
3	Ducks	
2	Frogs	
2	Ducks	
3	Frogs	
3	Ducks	
5	Pigs	
Requested	Animal	Given
10	Frogs	
8	Ducks	
8	Ducks	
10	Frogs	
6	Frogs	
10	Pigs	
6	Pigs	
8	Ducks	
6	Pigs	

Give-N Naturalistic

Date:

Pegs Requested	Given
5	
5	
2	
3	
2	

2	
3	
3	
5	
Pegs Requested	Given
10	
8	
8	
10	
6	
10	
6	
8	
6	

Give-N Role Play

Date:

Requested	Fruit	Given
5	Potatoes	
5	Carrots	
2	Tomatoes	
3	Potatoes	
2	Tomatoes	
2	Tomatoes	
3	Carrots	
3	Carrots	
5	Potatoes	

Requested	Fruit	Given
10	Carrots	
8	Carrots	
8	Tomatoes	
10	Potatoes	
6	Tomatoes	
10	Potatoes	
6	Potatoes	
8	Tomatoes	
6	Carrots	

WOC

Date:

Requested	Animal	Given
5	Pigs	
5	Frogs	
2	Pigs	
3	Ducks	
2	Frogs	
2	Ducks	
3	Frogs	
3	Ducks	
5	Pigs	

Requested	Animal	Given
10	Frogs	
8	Ducks	
8	Ducks	
10	Frogs	
6	Frogs	
10	Pigs	
6	Pigs	
8	Ducks	
6	Pigs	

Table S1

Mean age of children spilt by nursery setting

Nursery	M_{age} (months)	SD_{age}
A	44.4	6.84
B	42.9	8.28
C	44.2	6.20
D	45.1	6.67

Figure S1

Example card used for What's on the Card task

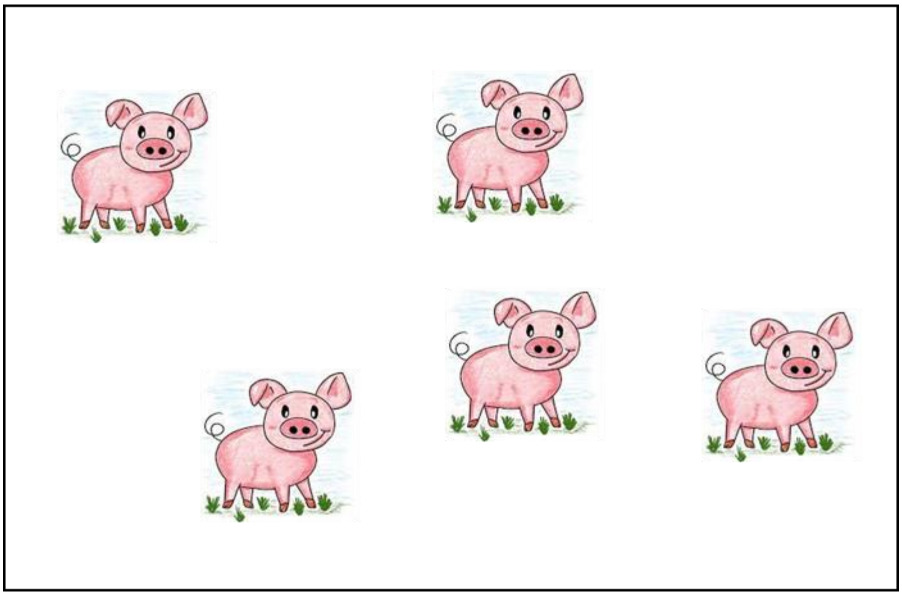

Table S2

Range of cards used for What's on the Card task

Number of animals on the card	Type of animal
Demo card	Pig, duck, frog
3	Pigs
3	Ducks
3	Frogs
4	Pigs
4	Ducks

4	Frogs
5	Pigs
5	Ducks
5	Frogs
6	Pigs
6	Ducks
6	Frogs
7	Pigs
7	Ducks
7	Frogs

Table S3

Binary regression model showing the effects of age, order of presentation and nursery setting at the first trial

	B	*SE*	*Z*
Intercept	-15.12	5.06	-2.99
Latin order 2	1.90	1.18	1.61
Latin order 3	1.42	1.07	1.33
Nursery D	-1.90	1.14	-1.67
Nursery A	-.03	1.26	-.02
Nursery B	2.13	1.17	1.82
Age	.29	.10	2.87**
	Model Fit		
	Nagelkerke R^2	LL	χ^2
	.52	-12.17	24.33***

R model equation: cardinal principle knowledge ~ latin + nursery + age

*p<.05 ***p<.001

Table S4

Binary logistic regression model showing the effects of age, order of presentation and nursery setting on the last trial

	B	*SE*	*Z*
Intercept	-9.60	3.47	-2.77
Latin order 2	.57	.98	.58
Latin order 3	.48	.85	.56
Nursery D	-.70	1.08	-.65
Nursery A	2.12	1.21	1.75
Nursery B	1.39	1.05	1.32
Age	.17	.07	2.47*
		Model Fit	
	Nagelkerke R^2	LL	χ^2
	.39	-8.29	16.58*
R model equation: cardinal principle knowledge ~ latin + nursery + age			

*p<.05 ***p<.001

Experimental Chapter 3: The Conceptual Development of Cardinal Proficiency

The previous chapters have explored the robustness of the Give-N task as a measure of cardinal proficiency. Here we found that different implementation methods affect children's displays of cardinal proficiency, and this is something current accounts of cardinal proficiency development do not account for. Furthermore, within the second chapter the findings question a central tenet of the knower-levels theory of cardinal proficiency development.

This central theoretical position suggests that a conceptual change occurs when children are able to give five items, and it is at this exact point that children understand the cardinal principle (Lee & Sarnecka, 2010, 2011; Le Corre et al., 2006; Sarnecka & Lee, 2009). As evidence for this conceptual advancement, they posit that children generalise their number knowledge, and their displays of cardinal proficiency should be comparable across all cardinal proficiency tasks. This was not supported by our results from chapter two.

Overall, the first two chapters begin to suggest that current theories of cardinal proficiency development, particularly the primary knower-levels theory, are not sufficient and do not provide a complete account of how this important skill develops. Likewise, the central tenets of this theory need further exploration. Do all children undergo a conceptual change when they are able to give five items? Or do we see a more gradual development of cardinal principle knowledge, with individual differences in children's developmental trajectories?

The third chapter within this thesis examines the central tenet of generalisation and explores the period of conceptual change. Assessing strategies adopted by children during cardinal proficiency tasks provides an alternative measure and perspective, enhancing our understanding of this conceptual change. We specifically examine the period when children

transition from being able to give four items to being able to give five items. It will explore whether, at this point, their knowledge generalises to all other numbers in their count list and other cardinal proficiency tasks. Additionally, we assess if, at this point, we see advancement in children's strategies adopted to pass cardinal proficiency tasks. This chapter directly corresponds to the second and third research questions within this thesis: *As suggested by the knower-levels account of cardinal proficiency development, do we see conceptual advancement in cardinal proficiency when children are able to give five items?* &
Can other measures of cardinal proficiency help to validate or advance the knower-levels account of cardinal principle development?

This paper is currently in preparation for publication:

Thomas, H., Lewis, C., & Dunn, K. (in preparation). A Micro-Genetic Exploration of Cardinal Proficiency Development: Sudden vs Gradual Change

A Micro-Genetic Exploration of Cardinal Proficiency Development: Sudden vs Gradual Change

Hannah Thomas

Charlie Lewis

Kirsty Dunn

Contributions

HT, KD & CL designed the study, with HT and CL writing the manuscript. HT recruited and children for the experiment, carried out the experiments, handled and coded the data, and performed all statistical analyses.

Abstract

Development of the cardinal principle is a key milestone in children's understanding of number. Here children understand the final number in a count represents the total number of items. Current theoretical accounts of cardinal proficiency development suggest that development of this key skill occurs abruptly, when children are able to give five items. Within this study we explore this assertion. By utilising a micro-genetic method and alternative measures of cardinal proficiency, such as strategy use and inaccuracy scores, we challenge the central tenets of current theory and explore the developmental pattern of this important skill. The findings do not support the central tenets of current theory and suggest that children's development of the cardinal principle is variable and is more likely to involve a gradual change. Importantly, we discuss the need for future micro-genetic studies, and the limitations with cross-sectional studies when exploring development of this key skill.

Introduction

Cardinal proficiency is a foundational skill in children's early number development. At the point when children understand that the final number in a count represents the total number of items in the set, they are credited with understanding the cardinal principle (Chu et al., 2016; Geary et al., 2018). This is a key landmark in children's number development, as once they understand this concept they are able to perform simple operations such as addition or subtraction. This crucial milestone is an important foundation for future mathematical achievement, and improving these early number skills contributes to a later mathematical advantage which is maintained into adulthood (Aunio & Niemivirta, 2010; Chu et al., 2015; Chu et al., 2016; Jordon et al., 2009; Koponen et al., 2013). It is therefore imperative that we have a reliable, evidence based, theoretical account of *how* cardinal proficiency develops. Recent evidence has raised questions for the dominant theoretical accounts of number development, which assert that the cardinal principle develops in a stepwise manner that suggests conceptual insight.

(Krajcsi & Fintor, 2023; Krajcsi & Reynvoet, 2023; Thomas et al., see chapter two). Questions have also been raised about the reliability of the primary measure for assessing this emergence (Thomas et al., see chapters one & two). As a result, the acquisition of the cardinal principle needs to be scrutinised, in detail, and different measures need to be explored to provide a secure evidence base for future theoretical accounts. We adopt a micro-genetic analysis of different measures of cardinal proficiency development to provide detailed, reliable evidence upon which the developmental pattern of cardinal proficiency can be explicated.

Theoretical challenges

The dominant, 'knower-levels' account of early number development suggests that children progress through incremental stages during their development of the cardinal principle (Lee & Sarnecka, 2010, 2011; Le Corre et al., 2006; Sarnecka & Lee, 2009; Wynn, 1990). When assessed using what has become the primary measure of cardinal proficiency, the Give-a-number or Give-N task[1], initially preschoolers are only able to give one item when asked, and at this stage they are deemed a 'one-knower'. Then a month or so later children progress to giving two items and become a 'two-knower', and so on, until they are able to give five or more items. According to the knower-levels account, at this point children undergo an abrupt conceptual change and now understand the cardinal principle – that the final number of a count represents the set size (Chu et al., 2016; Geary et al., 2018; Lee & Sarnecka, 2010, 2011; Le Corre et al., 2006; Sarnecka & Lee, 2009). Following this conceptual advance they quickly generalise this knowledge and are able to give all numbers of items within their count list, and their cardinal proficiency is equivalent across different cardinal principle tasks (Lee & Sarnecka, 2010, 2011; Le Corre et al., 2006; Sarnecka & Lee, 2009).

There have been recent challenges to this dominant account (Krajcsi & Fintor, 2023; Krajcsi & Reynvoet, 2023), mainly centring on the pattern of change. The knower-levels theory posits an abrupt development, once children are able to give five items (Lee & Sarnecka, 2010, 2011). However, Krajcsi & Fintor (2023), Krajcsi & Reynvoet (2023) and Thomas et al., (chapter two) suggest that the pattern of change may be more gradual. The findings from the first two studies by Krajcsi & Fintor and Krajcsi & Reynvoet revealed that some children were able to give numbers higher than five (e.g., seven or eight items) but not all the items in their count list, violating the generalisation tenet within the knower-levels theory. Secondly, these children showed similar performance on a comparison task to those who were unable to give five or more items. Specifically, sub-set knowers (those not able to give five or more) and children who were able to give six, seven or eight (but not all the

[1] See supplementary materials for a full description of the Give-a-number task

numbers within their count list) were only able to compare 'known' numbers i.e., numbers they were able to give during the Give-N task (Krajcsi & Reynvoet, 2023). Krajcsi and Reynvoet conclude that, given the lack of generalisation in their responses and their comparative performance matching that of sub-set knowers, there may be an additional category of children. Those able to give six, seven or eight etc., but not able to generalise may be larger number sub-set knowers, and these children are still conceptually sub-set knowers and not cardinal principle knowers. This challenges the generalisability claim within the knower-levels account of development. Krajcsi and Reynvoet suggest that not all children undergo a conceptual change when they are able to give five items, and the pattern of conceptual change may be more gradual as they progress through the knower levels.

A further challenge to the generalisability tenet is raised by Thomas et al (chapter two) who suggest that five, or more, knowers may not automatically generalise their cardinal principle knowledge within one test, like the Give-N, to other cardinal principle tasks. They showed that children's performance on the Give-N and What's on the Card task (a paper based cardinal principle task) were not equivalent even though they set out to assess the same ability. It may be that children's cardinal principle knowledge is not secure and may still be developing, with slight variations in procedure revealing this. As a result, Thomas et al.'s conclusions echo those of Krajcsi and Reynvoet (2023) and point to a more gradual pattern of change for cardinal principle development.

Furthermore, the primary measure of cardinal proficiency, the Give-N task, is inextricably linked to the knower-levels account of cardinal proficiency development, with the task and theoretical accounts formulated collectively. This confines theoretical assumptions and restricts the methods utilised to assess these. As such, it is timely to explore different ways of measuring cardinal proficiency.

With growing evidence challenging the generalisation tenet of the knower-levels account of development and suggesting that the pattern of change is more gradual than the abruptness posited by this account, it is important that we explore cardinal principle development in greater detail, and utilise alternative measures to validate current research further. Micro-genetic methods are often used to allow an in-depth analysis of the period of change and afford an alternative view to current numerical, categorical judgements of children's development.

Micro-genetic analysis

The present theoretical challenges, like those of Krajcsi & Reynvoet (2023) and Thomas et al., (chapter two), may be the product of the preponderance of cross-sectional studies, which may underestimate the complex nature of number development. Such a design can overlook the detail within children's performance across trials. Current accounts of cardinal proficiency development may be constrained by such methods. With cross-sectional studies dominating the literature, their suggestion of sudden conceptual change may be confounded due to their reliance upon single time-point analyses. In order to explore the pattern of change in detail, micro-genetic analysis is essential to test the stepwise account against one of gradual change. This method focuses on the period of development, taking repeated measurements frequently, trial-by-trial over this period, providing detailed quantitative and qualitative data about any key transitions (Siegler & Crowley, 1991). This analysis will hopefully allow a view of development where regressions and variability can be assessed, and overall development can be mapped By examining the transition process closely, this method is an effective way to elucidate *how* the change may occur and if this is stepwise or gradual (Flynn et al., 2006).

To our knowledge only one previous study has explored cardinal proficiency development using micro-genetic methods (Chetland & Fluck, 2007). Whilst the authors discuss the importance of understanding the process of development, their analysis has only two time points one week apart. The use of only two time points in order to measure the nature of developmental change has long been criticised (Rogosa, et al., 1982). Micro-genetic analysis must attempt to span the period of development, and the testing sessions must be frequent enough to gain an accurate understanding of the pattern of change (van Dijk & van Geert, 2007). Cardinal principle development is estimated to take a number of months (Sarnecka & Carey, 2008), therefore the analysis period and the two testing sessions within Chetland and Fluck's study are insufficient for a complete view of change.

Micro-genetic analysis and alternative measures of cardinal proficiency

The micro-genetic methodology also facilitates the use of alternative measures, allowing a shift away from 'knower-levels' with fixed, numerical, categorical judgements of children's development. Assessment of strategies that children adopt are an essential element of micro-genetic analysis, as these can help to explicate the timing of change as children transition between simple strategies to more advanced ones (van Dijk & van Geert, 2007). Is the transition gradual with variability in strategy use or is it abrupt? With regards to cardinal proficiency development it is hoped that such a study will inform our judgements regarding the pattern of change, be that gradual or abrupt. In relation to cardinal proficiency development, there are only two strategies reported within the literature. These are 'grabbing' items vs 'counting' items during the Give-N task (Chetland & Fluck, 2007). It is suggested that children transition from a simplistic grabbing strategy to a counting strategy when they are able to give five items and therefore, according to the knower-levels account of development, understand the cardinal principle (Le Corre et al., 2006, Sarnecka & Lee, 2009; Lee & Sarnecka, 2010, 2011; Wynn, 1990). However, the exact nature of children's strategies has not been explored in detail. Using micro-genetic analysis, therefore, will further validate such assertions and may lead to an alternative analysis. Likewise, further assessment of strategies should help to elucidate the pattern of developmental change.

In addition to analysing the possible strategies that children employ, assessment of the child's performance across repeated trials may facilitate our appraisal of developmental change. Whilst strategy assessment is customary within micro-genetic analysis, trial by trial analysis has not, to our knowledge, been explored. In the Give-N task children are required to show an understanding of a cardinal number by succeeding in giving that number two times out of three. Some sail through all trials while others err, sometimes at each successive level. As a result, we computed a measure of the degree to which children fail to understand trials, which we term 'inaccuracy'. Such an assessment has not been utilised before within cardinal proficiency research. Inaccuracy scores are computed as the difference between the number of items requested and the number of items children give during the Give-N task. These should provide a detailed span of data from which we can assess development, as opposed to knower-levels that place children into ranked categories based on dichotomous pass / fail judgements. Reflecting on the gradual view of development suggested by Krajcsi and Fintor (2023), Krajcsi and Reynvoet (2023) and Thomas et al.,(chapter two), accuracy scores will also provide a continuous, more detailed, measure without the confines of the categorical knower-levels approach, from which to analyse this alternative, gradual suggestion. Furthermore, by using micro-genetic methods, strategy analysis and inaccuracy measurements may be more reliable, as these accommodate variability, more so than static state, categorical measurements afforded from current methodologies.

Purpose of the current study

Given that questions have been raised regarding current theoretical assertions relating to the pattern of cardinal proficiency development, we aim to examine development of this crucial skill adopting a micro-genetic method and assessing development using alternative measures, inaccuracy scores and strategy use. Specifically, we aim to answer the following questions: Do children generalise cardinal proficiency abilities to all numbers in their count list and to other cardinal principle tasks once they are able to give five items? Is there variability in children's displays of cardinal proficiency over time? And subsequently, does the use of alternative measures support a gradual view of cardinal proficiency development? When responding to the questions we hope to advance current theoretical positions and possibly provide alternative, more reliable, measures of cardinal proficiency development. Overall, findings from this study should not only inform theory, but improve research practice throughout the developmental literature.

Hypotheses

All hypotheses are driven by evidence suggesting that the generalisation tenet within the knower-levels theory of cardinal proficiency may not be accurate and, once children are able to give five items during the Give-N task, they will not automatically be able to give all the numbers of items within their count list or generalise this knowledge to other cardinal proficiency tasks (Baroody & Lai, 2022; Baroody et al., 2023; Krajcsi & Fintor, 2023; Krajcsi & Reynvoet, 2023; Thomas et al., chapter two; O'Rear et al., 2024). Furthermore, the hypotheses are derived from corresponding, emerging evidence suggesting that cardinal proficiency may be a gradual developmental change, as opposed a sudden conceptual change once children are able to give five items, as suggested by the knower-levels account of development (Krajcsi & Fintor, 2023; Krajcsi & Reynvoet, 2023; Thomas et al., chapter two).

Hypothesis one

Based on the above evidence, it is predicted that once children are able to give five items during the Give-N task they will not automatically be able to give all the numbers of items within their count list. Alternatively, based on the knower levels theory, it is predicted that once children are able to give five items during the Give-N task they will automatically be able to give all the numbers of items within their count list.

Hypothesis two

Again, in accordance with the above evidence it is predicted that no noticeable advancement to a counting strategy will be a pre-cursor to children being able to give, specifically, five items during the Give-N task. Alternatively, in accordance with the knower levels account, it is predicted that a sudden advancement to a counting strategy will be a pre-cursor to children being able to give five items during the Give-N task.

Hypothesis three

In line with the above evidence, it is predicted that children's knower-levels, calculated from the Give-N task, would not be comparable with those calculated from the How Many and What's on the Card tasks. Alternatively, following an assumption within the knower-levels account of development, it is predicted that children's knower-levels calculated from the Give-N task will be comparable with those calculated from the How Many and What's on the Card tasks over the testing sessions.

Hypothesis four

From the perspective proposed above, it is precited that inaccuracy scores will provide a less variable (and therefore more reliable) measure of cardinal proficiency development, and these scores will gradually improve over the testing sessions.

Method

Participants

Being able to recite the number sequence up to ten and being classified as a three or four knower during the Give-N task were a pre-requisite for participation. As such, we initially screened 74 children from local nurseries within the Lancaster area. From these 13 children were eligible. However, we established a criterion for inclusion that children must participate in four or more sessions to be included. Participant number two failed to meet this criterion, only undertaking two testing sessions. The remaining 12 children were aged between 25 and 47 months (M_{age} = 45.07 months, SD_{age} = 5.77, 6 girls). See procedure section for testing session detail.

Materials

Traditional Give-N task

For this task a medium sized cuddly toy rabbit, named Rachel, and a plastic plate / tub were used for children to place different animals on and 'give' these to Rachel. The animals comprised three sets of small plastic animals; 15 frogs, 15 pigs and 15 ducks. Each set of animals was contained in a separate plastic tray / basket.

What's on the Card task

A set of A4 laminated cardboard cards were used for this task. Printed on the cards were different sets of animals, for example one card displayed five pigs (see Figure S1). In total there were 15 cards displaying different sets of animals plus one demonstration card (see Table S1 for a full list of the different cards).

How Many task

An A4 set of laminated cardboard cards were used. Again, there were 16 cards in total, including one demonstration card. This time each card had a single line of animals printed across the middle, with different numbers of animals on each card (see Figure S2 for an example card, and Table S2 for a full list of the different cards).

Procedure

All the tasks were completed in a quiet room at the child's nursery or preschool, and were video-recorded for offline coding. All tasks were completed on an approximately weekly basis, for a period of three months (see Table 1 for exact testing schedule). The tasks were presented in a random order to each child, at each weekly session. All tasks were completed with children seated at a table, with the researcher opposite.

Table 1

Participant testing schedule

	Session									
ID	0	1	2	3	4	5	6	7	8	Nursery
1	15/05	22/05	30/05	06/06	13/06	20/06	30/06	07/07	14/07	A
3	15/05	22/05	09/06	16/06	23/06	30/06	07/07	14/07		A
4	15/05	22/05	30/05	06/06	13/06	20/06	30/06	14/07	21/07	A
5	16/05	22/05	30/05	06/06	13/06	11/07	18/07			A
6	23/05	07/06	16/06	21/06	28/06	05/07	13/07			B
7	23/05	07/06	14/06	21/06	28/06	05/07	14/07			B
8	23/05	09/06	14/06	23/06	30/06	05/07	13/07			B
9	23/05	07/06	14/06	21/06	28/06	17/07				B
10	24/05	14/06	21/06	28/06	10/07					B
11	24/05	14/06	21/06	05/07	13/07					B
12	01/06	20/06	27/06	05/07	12/07	19/07				C
13	26/05	20/06	27/06	05/07	12/07	19/07				C

Give-N task

On the table in front of the researcher was Rachel the rabbit and the plastic tub. The three baskets of animals were placed next to the child on the table. The children were introduced to Rachel and were told “today you’re going to give her animals to play with”. After this, children received one practice trial, where they were asked to give Rachel one pig and to put it on her plate. Following the practice trial children received feedback; either “ooh lovely, thank you” if they gave the correct number of pigs or, if they gave the incorrect number of pigs “oh I don’t think that’s just one pig, Rachel only wants one pig. Can you give her just one pig please?" If the child did not respond the trial was repeated in an encouraging way, and the test trials followed immediately.

During the test trials children were asked to give Rachel the rabbit different numbers of animals. Numbers between three and seven (inclusive) were asked for in an incremental order, and adopting a titration method using two, possibly three, trials for each number requested. Unlike the standard titration method children were asked for all numbers between three and seven, regardless of their correct or incorrect responses to lower numbers requested. The animals were returned to their specific baskets after each trial to maintain the same set size from which children selected the animals.

Children were given one prompt per testing session. Upon an incorrect response, children were told "hang on, that's not 'x', let's try again." For all other responses / trials children were given neutral feedback, "ooh, thank you", after passing the animals.

What's on the Card task

Following a similar procedure to Sarnecka & Negen, 2019 (https://osf.io/eznht/) the task began with the researcher showing the child the demonstration card, and saying "look here are some cards with pictures of frogs, pigs and ducks on". The practice trial followed with the researcher showing the child a card with one pig on and asking "what's on this card?". If the child replied with "a pig" the researcher explained that in this game we use number words, so we say "one pig", the child was asked to repeat this. The test trials followed with the researcher showing the child the cards individually, and asking "what's on this card?". The incremental order the cards where shown to the child mirrored the presentation during the Give-N task. Likewise, feedback during the test phase mirrored that of the Give-N task.

How Many task

The procedure mirrored that of the What's on the Card task. However, children were asked "How many pigs (or frogs / ducks) are there?" rather than "what's on this card?".

Scoring

Numbers between three and seven (inclusive) were asked for in an incremental order, adopting a titration method. Unlike the standard titration method children were asked for all numbers between three and seven, regardless of their correct or incorrect responses to lower numbers requested. Therefore, children were assigned a knower-level score based on the highest number they were able to give two out of three times. As such, this method did not stipulate that children must have given numbers lower than their knower-level category accurately.

Inaccuracy scores are the difference between the number of items requested minus the number of items given by the child on each trial.

Coding

Table 2 details the different strategies children used during the Give-N task. These were derived from the initial coding videos and coding videos from Thomas et al., (chapter two).

Table 2

Strategy definitions

Strategy label	Strategy definition
Grab	Grabs and then dumps items into the tray
Count	Individual counting out of items (aloud or silent)
Grab & Count	Combination of A & B
Grab same number	Grabs and dumps same number of items each time

Inter-rater reliability

Offline video coding was completed by the primary researcher, with a random sample of four children second coded by a trained research assistant. The second coding produced an inter-rater reliability of 93%. The discrepancies were resolved by discussion.

Results

Assessment of sudden conceptual change

In response to hypotheses one, two and three, and in order to assess if cardinal proficiency development is a sudden conceptual change once children are able to give five items, we examined if children automatically generalise knowledge of the cardinal principle when they are to able to give five items. Specifically, once children give five items are they able to give the correct number of items for all the numbers within their known count list, and do we see a sudden advancement to a counting strategy to achieve these results. Furthermore, once children are able to give five items do their scores on equivalent tasks mirror this knowledge. In order to assess this, we analysed children's knower-levels and the strategies they used during the Give-N task over the testing period and compared their Give-N knower-levels to those on alternative cardinal proficiency tasks.

Furthermore, we analysed children's variability of knowledge within the period of change, by assessing the knower-levels across the testing sessions.

Do Children Generalise Knowledge to The Rest of Their Count List?

Knower-Levels Assessment. Assessment of the children's knower-levels over the testing period revealed highly variable results. All the children showed progressions and regressions over the three-month period, suggesting their understanding of number is dynamic. Figure 1 displays the knower levels for each of the twelve participants. The dots represent the child's knower level at each of the testing sessions – note that some attended all eight, but others dropped out after week 4. The regression slopes plot the linear trend of performance taking all trials into account. Participants six, ten, eleven and twelve all show regressions lasting a number of weeks, suggesting that the regressions are developmental patterns rather than extraneous data due to behavioural noise. Whereas, participants one, five, eight and thirteen all show improvement in performance across the weeks

When specifically assessing the children's grasp of cardinality, using the usual benchmark of selecting five items during the Give-N task, the findings suggest that their knowledge is inconsistent and fluctuates from session to session. Nine out of the twelve children were able to give five or more items at one or more time point/s across the sessions, but not consistently. Additionally, the three children (participants one, five & thirteen) who were able to give five (or more) on four (or more) occasions still show regressions during at least one testing session.

Importantly, and in contrast to the hypothesis deriving from the knower levels account, once children are able to give five items, they do not appear to generalise this knowledge and are unable to give all numbers of items within their count list consistently. To test this observation of the data a binomial logistic regression model was constructed using the GLM function within the R base package. Whether the child was able to give five items, and therefore a cardinal principle knower (as per the knower-levels theory of development) was entered as the binary predictor variable (yes / no), and whether the child was also able to give seven items and therefore classified as a seven knower (yes / no) was entered as the outcome variable. This seemed to confirm that being a cardinal principle knower at any session does not predict being a seven knower during subsequent sessions (R_2 =.37, β = 20.29, p = .99). Therefore, we do not find support for a sudden conceptual development of the cardinal principle.

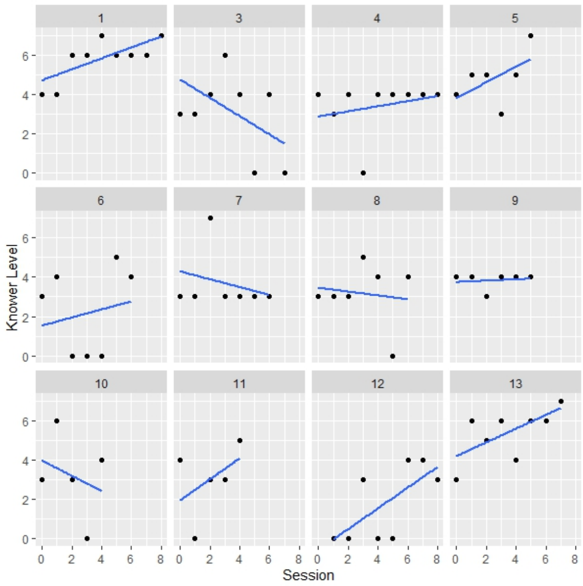

Figure 1

Knower levels for each participant across each session

Strategy Assessment. We devised four clear strategies based on each trial: Grabbing any number of items, without obvious counting (Grab); counting individual items (Count); grabbing a number of items, then counting individual items to add to the set (Grab & Count); always grabbing the same number i.e., that could be depicted as 'systematic grabbing' (Grab same number).

The subsequent analysis of these strategies revealed no sudden advancement to a counting strategy prior to or when children were able to give five items (see Figure 2). This, therefore, suggests that development of the cardinal principle is not a sudden conceptual change after which children are able to give five items. A binomial logistic regression model was constructed using the GLM function within the R base package, with cardinal principle knower (yes / no) as the outcome variable and use of a counting strategy as a predictor variable. The use of a counting strategy did predict being a cardinal principle knower (R_2 = *.07, β* = 1.46*, p* = .016). However, these results need to be interpreted with caution as a large majority the children used a counting strategy across many trials (see below and Figure 2).

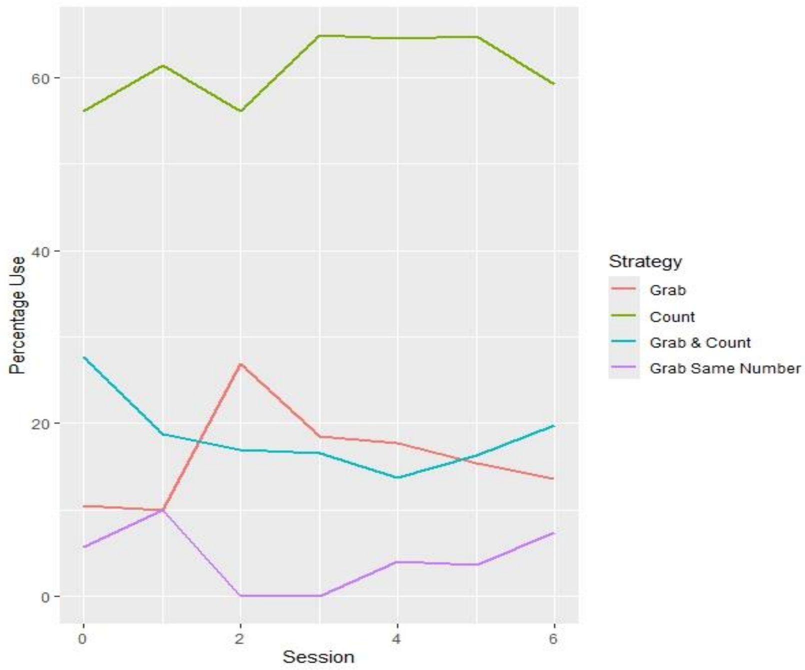

Figure 2

Percentage strategy use across the sessions

*Sessions 7 & 8 have been omitted from the graph as these final testing sessions only consisted of a small number of children

Figure 2 shows that the counting strategy was dominant throughout the sessions, regardless of the number of items requested and the nursery setting the children attended. Seven out of the twelve children adopted the counting strategy consistently, with a counting strategy used 59% of the time, a grab and count strategy 19% of the time, a pure grab strategy 17% of the time and a grab the same number strategy 5% of the time. See supplementary materials for additional statistical breakdown.

Do Children Generalise Knowledge to Other Tasks?

We assessed children's grasp of cardinality using the benchmark of being able to identify five items during the What's on the Card and How Many task. Again, these tasks reveal variability in children's understanding, with little generalisation of knowledge when children are able to identify five items (see Figure 3). Spearman's correlations between knower-levels on the Give-N task, knower-levels on what's on the card task and knower-levels on the how many task revealed no significant relationships between the tasks at any time point (see Table 3). Closer inspection did suggest that despite the lack of significance knower-levels across times one and two produced high r_s values when correlated with what's on the card knower-levels at time one and two. We therefore pooled the data and undertook ordinal regressions with Give-N knower-levels as the outcome variable and the concurrent What's on the Card and How Many knower-levels as predictor variables. The results suggested that Give-N knower-levels at time seven correspond with what's on the card knower-levels at time seven, and Give-N knower-levels correspond with how many knower-levels at time three, five and seven (see Table 4). With only three out of the seven sessions producing any corresponding knower-levels and no significant correlations at any sessions or at any knower-level (including children who gave five or more), the results suggest that children do not automatically generalise cardinal principle, or knower-level, knowledge to other tasks. Again, this challenges the generalisability tenet within the knower-levels account of development, and therefore suggests that development of the cardinal principle is not a sudden conceptual change when children are able to give five items. What we see is a feature of children's number development and this is variable and sensitive to task demands.

Furthermore, the high variability present across all three tasks suggests that this pattern is not a feature only of the Give-N task, but rather a reflection of children's grasp of number during this development. Recognition of this is important in our assessment of knower levels as most studies have used cross-sectional tests, and the assumption that these are a true reflection of children's knowledge may not be warranted.

Follow up What's on the Card analysis. As study two found that children performed at a higher level on the What's on the Card task than the Give-N procedure, we replicated the analysis for each week in study three. To make the analyses comparable we checked for normality and conducted a within participants t test at each week of testing over the first six sessions (we used parametric analysis in study two). In only one of weeks (week 4) was the effect partly replicated ($t(11) = 2.08$, $p = .031$ [1 tailed], $p = .062$ [2 tailed]). It is the case that in four of the six sessions (2-5), children's knower levels were higher in the What's on the Card task but, as Figure 3 suggests, in this repeated testing regime the evidence of one test being easier than the other was not as clear as in the larger sample tested only once in study two (see Table S4 for non-significant results).

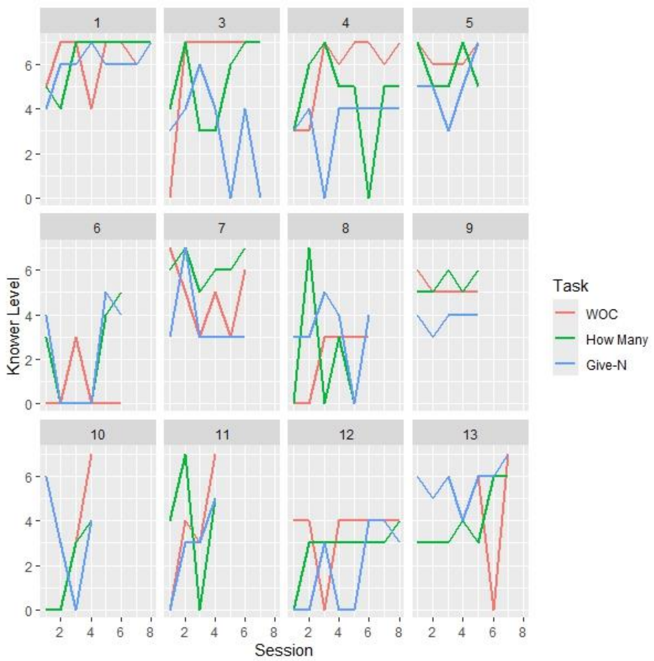

Figure 3

Knower levels assessed on the what's on the card, how many and give-n task across the testing sessions split by participant

Table 3

Correlation matrix showing relationships between knower-levels on the give-n, how many and what's on the card tasks

Task	*n*	Give-N Knower-Level Session					
How Many Knower-Level		1	2	3	4	5	6
Session 1	12	.209					
Session 2	12		.452				
Session 3	12			.089			
Session 4	12				.399		
Session 5	10					.198	
Session 6	8						.085
What's on the Card Knower-Level							
Session 1	12	.504					
Session 2	12		.563				
Session 3	12			.260			
Session 4	12				.399		
Session 5	10					.327	
Session 6	8						-.092

Table 4

Ordinal regression models showing give-n knower levels predicted by how many and what's on the card knower-levels

How Many predictor model			
	B	*SE*	*P*
Session three	1.54	.78	.049*
Session four	1.53	.87	.079
Session five	1.89	.76	.014*
Session six	1.49	1.73	.083
Session seven	2.61	.81	.001*
	Model Fit		
	Nagelkerke R^2	LL	χ^2
	.14	-6.31	12.62*
Model equation: Give-N knower-level ~ how many knower-level			

What's on the card predictor model			
	B	*SE*	*P*

Session three	.36	.72	.62
Session four	-.48	.80	.54
Session five	.94	.83	.26
Session six	1.41	.78	.07
Session seven	1.72	.68	.012*
		Model Fit	
	Nagelkerke R^2	LL	χ^2
	.16	-7.03	14.07*
Model equation: Give-N knower-level ~ what's on the card knower-level			

*p<.05

Assessment of the variability within development

To assess the developmental variability in children's knower-levels over the testing sessions we explored each session as a predictor for subsequent sessions. We started by looking at connections between each test and the next. For example, we investigated whether performance at the baseline (week 0) predicted that in the next session (week 1), then from week one to week two, and so-on. In each analysis we loaded the child's knower level and inaccuracy score[3] the week before as explanatory variables. The sample was complete until week four, after which there were missing data due to child absences.

In the first set of analyses, none of the models that we constructed, examining weeks 1-6 as dependent measures, were significant. We then examined the same models with a two or three-week latency. Here some were significant models. These suggested that knower-levels at time two and time three were predicted by inaccuracy scores at the baseline testing session (week 0) (see Table 4, panels A and B). Knower-levels at time two were predicted by knower-levels at this initial testing session (see Table 4, panel B). Knower-levels at time four were predicted by knower-levels on the second testing session (see Table 4, panel C) and

[3] Please see following paragraph for explanation of inaccuracy scores.

knower-levels at time five were predicted by knower-levels at the first testing session (see Table 5, panel D). See supplementary materials for the non-significant models.

Table 5

Significant ordinal regression models showing the knower-levels predicted by previous sessions.

Panel A			
	B	*SE*	*P*
Session zero knower-level	3.80	1.84	.041
Session zero inaccuracy	.248	.11	.022
		Model Fit	
	Nagelkerke R^2	LL	χ^2
	.52	30.45	8.40*
Model equation: knower-level session two ~ knower-level session zero + inaccuracy score session zero			
Panel B			
	B	*SE*	*P*
Session zero inaccuracy	.17	.08	.035
		Model Fit	
	Nagelkerke R^2	LL	χ^2
	.40	26.91	5.68*
Model equation: knower-level session three ~ inaccuracy score session zero			
Panel C			
	B	*SE*	*P*
Session two knower-level	.82	.35	.020
		Model Fit	
	Nagelkerke R^2	LL	χ^2
	.37	23.35	5.08
Model equation: knower-level session four ~ knower-level session two			
Panel D			
	B	*SE*	*P*
Session one knower-level	2.78	1.26	.027
		Model Fit	
	Nagelkerke R^2	LL	χ^2
	0.75	20.89	13.03**

Model equation: knower-level session five ~ knower-level session one

Assessment of gradual conceptual change

In order to examine if a gradual pattern of change is more representative than sudden conceptual development of the cardinal principle, thus testing hypothesis four, we created an alternative measure. Rather than use knower-levels, which by nature are categorical, we devised an inaccuracy score measure. This was calculated as the difference between the number requested and the number given by the children during the Give-N task. We assumed that this might allow us to identify a linear pattern of development should one be in evidence, and may provide a more precise, detailed measure or performance on the Give-N task.

This assessment revealed that children's inaccuracy scores in the Give-N task decrease gradually over the sessions (see Figure 4 for analysis of the data pooled across participants). This applied to all the numbers requested despite children's levels of number knowledge (see Figure 5 for an analysis of the six children who completed all the trials). These patterns suggests that, although number development may be complex, we do see gradual over improvement in performance in individual trials over time and, given the mixed findings of knower level scores over time, supports the assertion that development of the cardinal principle may be gradual, as opposed to sudden.

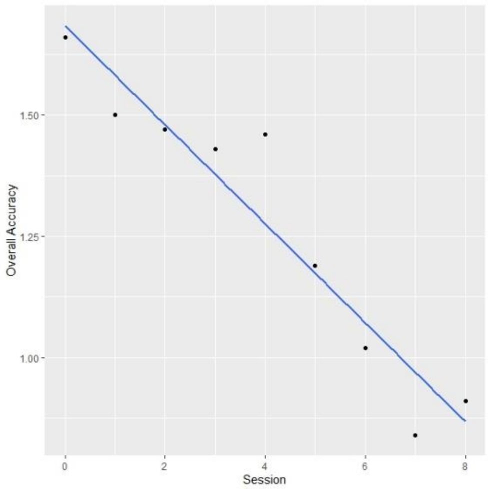

Figure 4

Overall inaccuracy scores per testing session

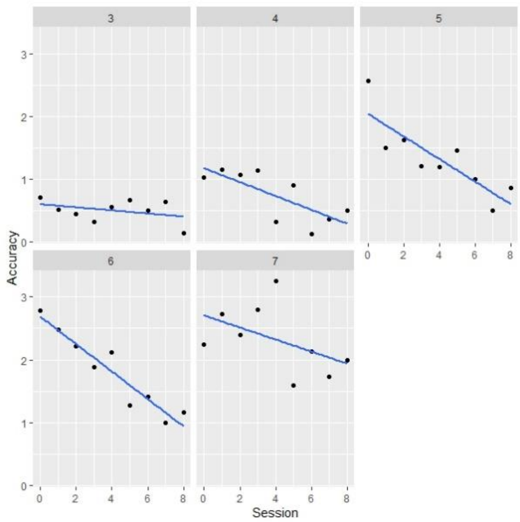

Figure 5

Overall inaccuracy scores for each number requested across all testing sessions

Given the consistency of change over trials across children, the revised inaccuracy measure may provide more consistent results across tasks. However, this assertion needs further analysis with a larger sample size. Nevertheless, the linear regression slopes in Figure 5 show only the general trends in inaccuracy across trials. The dots in that figure show variability in performance. Figure 6 plots the inaccuracy scores for all three tests in each child, across sessions. As with the knower levels data, this shows large individual variability in the numbers of incorrect responses in each session.

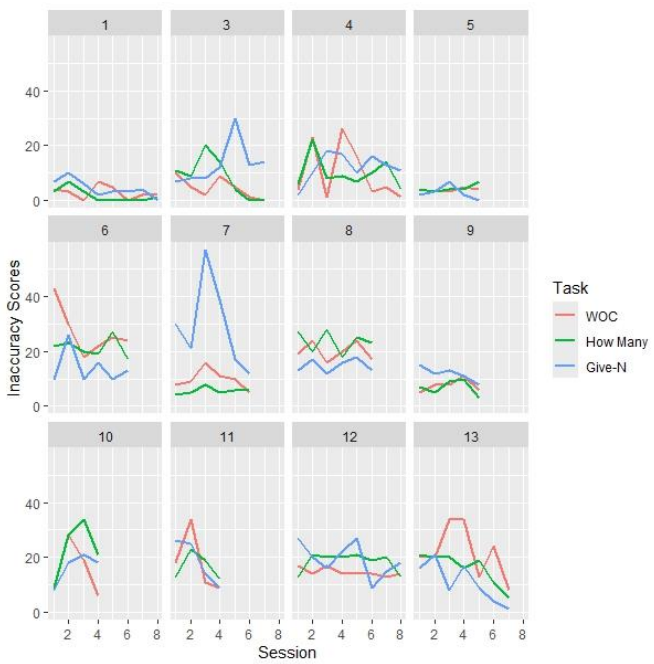

Figure 6

Overall inaccuracy scores for each number requested across all testing sessions

Discussion

The current study has highlighted the variability in children's knower-levels over time, and the lack of conceptual generalisation of the cardinal principle across tasks and when children are able to give five items. Once children were able to give five items (and would typically be deemed cardinal principle knowers), they were unable to consistently give the correct number of items for all numbers within their count list. Likewise, the knower-levels assigned to children during the standard Give-N task lacked comparability with those assigned during the What's on the Card and How Many tasks. In addition, the current study has revealed that the strategies children utilised during the Give-N task are consistent over time, with no sudden change of approach once children are able to give five items. It was apparent that a counting strategy was dominant for a large proportion of children irrespective of their success in the Give-N task. Furthermore, the results suggest that inaccuracy scores may provide an alternative, more reliable measure of cardinal proficiency development. As inaccuracy scores are more comparable across tasks and appear to have a slightly reduced variability in comparison to the knower-levels measures. Finally, our results highlight the higher variability within children's number development and strongly suggest that the 'level' identified in the dominant cross-sectional studies within cardinal proficiency research may be open to question. As children's ability varied from session to session, it is apparent that single timepoint studies may not reflect their true abilities.

Knower-levels

With regards to the knower-levels measure, both the hypotheses deriving from the knower levels theory were not supported. Our results suggest that, when using two trials out of three at a single time point as a measure of cardinal proficiency, the number identified is not consistent across related cardinal proficiency tasks. Likewise, we did not find a sudden advancement in children's displays of cardinal proficiency once they were able to give five items.

The lack of comparability across tasks could be explained by procedural demands of each. The Give-N task may be procedurally more demanding than the paper based How Many or What's on the Card tasks. 'Give-N' requires children to hold a number within their working memory, whilst then manipulating objects and counting them out. As such, these additional task demands may mask children's displays of cardinal proficiency. This argument has been posited a number of times within the literature (Baroody & Lai, 2022; Baroody et al., 2023; O'Rear et al., 2024). However, the results here suggest that no one task is easier or harder for children. What we do see is high levels of variability in children's knower-levels over time, regardless of the task.

Previous discussions regarding task difficulty have focused on task demands and have tried to separate elements of cardinal proficiency (Baroody & Lai, 2022; Baroody et al., 2023; O'Rear et al., 2024). The nature of these studies may be constrained by the use of cross-sectional design and adherence to the knower-levels theory of development. Our micro-genetic design has revealed the variability of children's assigned knower-levels over time. This suggests that perhaps our focus should move away from task demands and we should focus on examining the nature and extent of variability to provide a deeper level of understanding when assessing children's number development.

Indeed, further examination of this variability did reveal that children's displays of number knowledge may show a variable, but gradually changing pattern of development with their number knowledge at later sessions predicted by their knowledge two or three weeks earlier. This may suggest that children's insecure knowledge during the period of development produces variable results as they wrangle with conceptual understanding. In line with a dynamic systems approach children's behaviours are emergent and develop from problem solving during tasks, and development is a product of children's efforts over time. Systems lose stability in order to develop, therefore we should expect to see variability in performance whist the system stabilises and gradually advances (Thelan, 2005). We suggest that children are gradually acquiring conceptual knowledge as they progress through knower-levels, reaching a point of secure understanding occurring at different times for different children. This suggests a long period of wrestling with ideas identified in the fluid nature of performance across weeks and gradual acquisition of knowledge corresponding to and supporting the alternative and more gradual account of cardinal proficiency development posited by Krajcsi and Finton (2023) and Krajcsi and Reynovet (2023).

In addition, the lack of conceptual generalisation (to other tasks and to known numbers) when children are able to give five items, is problematic to the knower-levels account of development. If our findings are replicated, this will require a re-evaluation of this dominant theoretical account. Our findings question central tenets of this theory, and the inextricably linked, principle measure. Our results suggest that it is appropriate to consider this different view of cardinal proficiency development, like Krajcsi's gradualist account, and correspondingly consider different measures of proficiency.

Strategy use

With regards strategy use, again, the hypothesis deriving from the literature was not supported. Examination of the strategies children use during the Give-N task revealed no advancement in the strategies children use just before, or just after children are able to give five items. This again raises questions regarding the assertions within the knower-levels account of cardinal proficiency. Their suggestion that a sudden conceptual change occurs when children are able to give five items (Lee & Sarnecka, 2010, 2011; Le Corre et al., 2006; Sarnecka & Lee, 2009; Wynn, 1990), should have produced results where an advancement in strategies used by children is noted at this point. Instead, our results suggest that children have a primary strategy they use during the Give-N task, and this is consistent over time. Furthermore, a counting strategy is dominant with the majority of children, regardless of their knower-level. This suggests that children's knowledge may not underpin the strategies they utilise during tasks, rather the instruction provided by early years settings may drive strategy use which may in turn facilitate development. With no sudden change in strategy use when children are able to give five items, our results may reflect a gradual gaining of knowledge through instruction, experience and maturation. However, additional research would be required to explicate this relationship.

Inaccuracy scores

The results from this study tentatively support our inaccuracy score hypotheses. We find that plots of linear change over the trials and analysis of individual scores suggest that inaccuracy may be less variable over time and that change occurs. This finding corresponds with the gradual view of cardinal proficiency (Krajcsi & Fintor, 2023; Krajcsi & Reynovet, 2023), and once again challenge a central assertion of the knower-levels account of cardinal proficiency development. There seem to be no findings within a micro-genetic analysis which suggest that children undergo a sudden conceptual development when they are able to give five items (Lee & Sarnecka, 2010; Sarnecka & Carey, 2008; Sarnecka & Lee, 2009). Development may be more individual than suggested by the knower-levels account.

The data presented here suggest that by removing the confines of current measures and theoretical assertions, we may be more able to assess the true pattern of development in its complexity. Inaccuracy scores appear to be slightly more stable over time, which may be due to the continuous, unbounded nature of this measure, as opposed to the pass-fail categorical judgements afforded from knower-levels. Use of inaccuracy scores as an additional or alternative measure of cardinal proficiency in future studies may enhance our theoretical understanding of *how* cardinal proficiency develops. This measure may prove more reliable, and adoption of this measure facilitates further exploration of a gradual view of cardinal proficiency development.

Conclusion

The adopted micro-genetic methodology has allowed an alternative view of cardinal proficiency development. From this we have highlighted that children's developing knowledge of the cardinal principle is highly variable, and their routes to understanding cardinality may be more complex than current theoretical accounts acknowledge. Likewise, micro-genetic exploration has allowed us to explore how children approach cardinal principle tasks by assessing the strategies they use, providing an alternative measure of development. Furthermore, we have been able to explore, in detail, the pattern of change as children develop their knowledge. Our data support a gradual view of cardinal proficiency development.

Whilst the micro-genetic method has provided a number of insights it is acknowledged that this method is time consuming and time critical, and the current study, although over a number of months, suggest a needed to assess development over an even longer period. Capturing the period of change is an essential component of micro-genetic studies. However, this is difficult and cardinal proficiency development may need to be assessed regularly even over a twelve-month period to fully explicate the pattern of cardinal principle development. The current study has supported recent challenges to current theoretical accounts of development and highlighted the importance of exploring this skill over time. However, we acknowledge the limitations of time scales and sample size, and therefore assertions from these results are tentative.

Notwithstanding the above limitations we clearly advocate for use of micro-genetic methods when exploring development of the cardinal principle. It is evident that cross-sectional, single timepoint studies are limited in their assessment of children's abilities. Without assessment over time it is difficult to theorise, with confidence, how this important skill develops. Finally, we support and provide evidence for a variable, but gradual account of cardinal proficiency development.

References

Aunio, P., & Niemivirta, M. (2010). Prediciting children’s mathematcial performance in grade one by early numeracy. *Learning and Individual Differences, 20,* 427-435.

Baroody, A. J., & Lai, M. (2022). The development and assessment of counting based cardinal nummber concepts. *Educational Studies in Mathematics, 111*(2), 185-205.

Baroody, A. J., Mix, K. S., Kartal, G., & Lai, M. (2023). The development and assessment of early cardinal number concepts. *Journal of numerical cognition, 9*(1), https://doi.org/10.5964/jnc.10035.

Chetland, E., & Fluck, M. (2007). Children’s performance on the ‘give-x’ task: A microgenetic analysis of ‘counting’ and ‘grabbing’ behaviour. *Infant and Child Development, 16*(1), 35-51.

Chu, F. W., vanMarle, K., & Geary, D. C. (2015). Early numerical foundations of young children's mathematical development. *Journal of Experimental Child Psychology, 132*, 205-212.

Chu, F. W., vanMarle, K., & Geary, D. C. (2016). Predicting Children's Reading and Mathematics Achievement from Early Quantitative Knowledge and Domain-General Cognitive Abilities. *Frontiers in Psychology.* https://doi.org/10.3389/fpsyg.2016.00775 .

Flynn, E., Pine, K., & Lewis, C. (2006). The microgenetic method: Time for change. *Psychologist, 19* (3), 152-155.

Geary, D. C., vanMarle, K., Chu, F. W., Rouder, J., Hoard, M. K., & Nugent, L. (2018). Early Conceptual Understanding of Cardinality Predicts Superior School-Entry Number-System Knowledge. *Psychological Science, 29*(2), 191-205.

Jordon, N. C., Kaplan, D., Ramineni, C., & Locuniak, M. N. (2009). Early Math Matters: Kindergarten Number Competence and Later Mathematics Outcomes. *Developmental Psychology, 45*(3). 850-867.

Koponen, T., Salmi, P., Eklund, K., & Aro, T. (2013). Counting and RAN: Predictors of arithmetic calculation and reading fluency. *Journal of Education Psychology, 105*(1), 162-175.

Krajcsi, A., & Fintor, E. (2023). A refined description of initial symbolic number acquisition. *Cognitive Development, 65*, 101288. https://doi.org/10.1016/j.cogdev.2022.101288

Krajcsi, A., & Reynvoet, B. (2023). Miscategorised subset-knowers: Five and six knowers can compare only the number they know. *Developmental Science, 27*(1), e13430.

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*(2), 130-169.

Lee, M. D., & Sarnecka, B. W. (2010). A model of knower-level behaviour in number concept development. *Cognitive Science, 34,* 51–67.

Lee, M. D., & Sarnecka, B. W. (2011). Number-knower levels in young children: Insights from Bayesian modelling. *Cognition, 120,* 391-402.

O'Rear, C. D., Kirkland, P. K., & Purpura, D. J. (2024). The how many and give-n tasks: Conceptually distinct measures of the cardinality principle. *Early Childhood Research Quarterly, 66,* 61-74.

Rogosa, D., Brandt, D., & Zimowski, M. (1982). A growth curve approach to the measurement of change. *Psychological Bulletin*, *92*(3), 726

Sarnecka, B. W., & Carey, S. (2008). How counting represents number : What children must learn and when they learn it. *Cognition, 108*(3), 662-674.

Sarnecka, B. W., & Lee, M. D. (2009). Levels of number knowledge during early childhood. *Journal of Experimental Child Psychology, 103*, 325-337.

Sarnecka, B.W., & Negan, J. (2019). Longitudinal number-knower data. *Open Science Framework*, DOI 10.17605/OSF.IO/EZNHT

Siegler, R. S., & Crowley, K. (1991). The microgenetic method: A direct means for studying cognitive development. *American Psychologist, 46*(6), 606-620.

Thelan, E. (2005). Dynamic Systems Theory and the complexity of change. *Psychoanalytic Dialogues, 15*(2), 255-283.

Thomas, H. L., Lewis, C. & Dunn, K. (in preparation). A Systematic Review and Meta-analysis of the Give-N task as a measure of Cardinal Proficiency.

Thomas, H. L., Lewis, C., & Dunn, K. (chapter 2). Examining the Influence of Pretend Play on Children's Displays of Cardinal Proficiency.

van Dijk, M., & van Geert, P. (2007). Wobbles, humps and sudden jumps: A case study of continuity, discontinuity and variability in early language development. *Infant and Child Development, 16*, 7-33.

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

Supplementary materials

Figure S1

Example card used for What's on the Card task

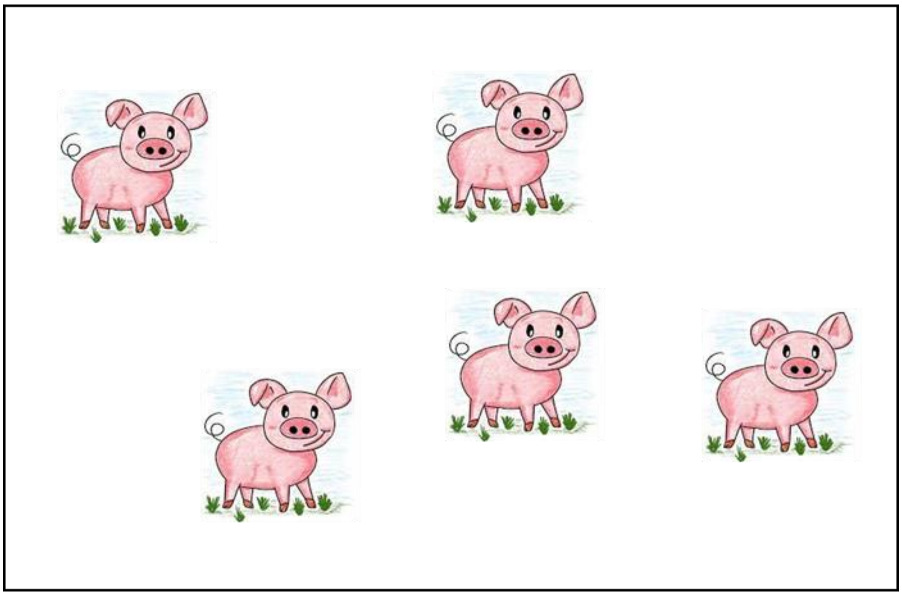

Figure S2

Example card used for How Many task

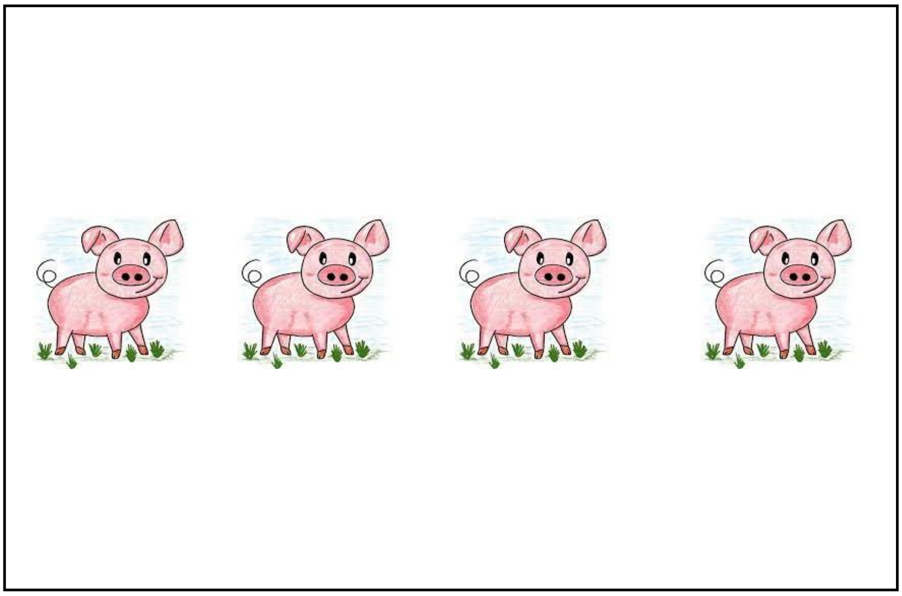

Table S1

Range of cards used for What's on the Card task

Number of animals on the card	Type of animal
Demo card	Pig, duck, frog
3	Pigs
3	Ducks
3	Frogs
4	Pigs
4	Ducks
4	Frogs
5	Pigs
5	Ducks
5	Frogs
6	Pigs
6	Ducks
6	Frogs
7	Pigs
7	Ducks
7	Frogs

Table S2

Range of cards used for the How Many task

Number of animals on the card	Type of animal
2	Ducks (demo card)
3	Pigs
3	Ducks
3	Frogs

4	Pigs
4	Ducks
4	Frogs
5	Pigs
5	Ducks
5	Frogs
6	Pigs
6	Ducks
6	Frogs
7	Pigs
7	Ducks
7	Frogs

Table S3

Non-significant ordinal regression models showing the knower-levels predicted by previous sessions.

Session Two (outcome)			
	B	*SE*	*p*
Session one knower-level	.40	.36	.27
Session one inaccuracy	.01	.07	.88
		Model Fit	
	Nagelkerke R^2	LL	χ^2
	.14	38.50	1.73
Model equation: knower-level session two ~ knower-level session one + inaccuracy score session one			
Session Three (outcome)			
	B	*SE*	*p*
Session two knower-level	.31	.30	.30
Session two inaccuracy	-.04	.08	.62
		Model Fit	

	Nagelkerke R^2	LL	χ^2
	.18	33.16	2.20
Model equation: knower-level session three ~ knower-level session two + inaccuracy score session two			
Session Three (outcome)			
	B	*SE*	*p*
Session one knower-level	.17	.35	.62
Session one inaccuracy	.03	.07	.65
		Model Fit	
	Nagelkerke R^2	LL	χ^2
	.02	35.11	.25
Model equation: knower-level session three ~ knower-level session one + inaccuracy score session one			
Session Four (outcome)			
	B	*SE*	*p*
Session three knower-level	.28	.27	.307
Session three inaccuracy	-.04	.04	.349
		Model Fit	
	Nagelkerke R^2	LL	χ^2
	.23	29.74	2.85
Model equation: knower-level session four ~ knower-level session three + inaccuracy score session three			
Session Five (outcome)			
	B	*SE*	*p*
Session zero knower-level	1.61	1.49	.28
Session zero inaccuracy	-.07	.08	.43
		Model Fit	
	Nagelkerke R^2	LL	χ^2
	.32	28.78	3.75
Model equation: knower-level session five ~ knower-level session zero + inaccuracy score session zero			
Session Five (outcome)			
	B	*SE*	*p*
Session two knower-level	.23	.29	.43
Session two inaccuracy	-.05	.09	.61
		Model Fit	

	Nagelkerke R^2	LL	χ^2
	.13	32.61	1.30
Model equation: knower-level session three ~ knower-level session two + inaccuracy score session two			
Session Five (outcome)			
	B	*SE*	*p*
Session three knower-level	-.09	.27	.75
Session three inaccuracy	-.03	.04	.46
		Model Fit	
	Nagelkerke R^2	LL	χ^2
	.09	31.67	.86
Model equation: knower-level session five ~ knower-level session three + inaccuracy score session three			
Session Five (outcome)			
	B	*SE*	*p*
Session four knower-level	2.2	.35	.53
Session four inaccuracy	-.07	.08	.915
		Model Fit	
	Nagelkerke R^2	LL	χ^2
	.27	29.49	3.04
Model equation: knower-level session five ~ knower-level session four + inaccuracy score session four			

Table S4

Results from t tests assessing performance on the What's on the card task in comparison to the Give-N task across each session

Session	*M*	*SD*	*T*	*df*
1	.91	2.59	1.17	10
2	-.03	2.31	-.13	11
3	-.92	2.86	-1.11	11
4	-1.17	1.95	-2.08*	11
5	-1.30	3.23	-1.27	9
6	.13	3.40	.10	7

* significant one-tailed test $p<.05$

Summary of individual results

ID1: Is able to give numbers higher than five at session two. Uses a dominant B strategy. Accuracy improves for all numbers across all sessions.

ID3: Is able to give six items at session two, but this then drops below five at subsequent sessions. Uses a dominant C strategy. Accuracy improves for all numbers across most sessions (excluding session three).

ID4: Unable to give more than four items. Uses a dominant A strategy. Accuracy improves for numbers lower than six but remains poor when asked for six or seven items.

ID5: Is able to give five items at session one and is then able to give seven items consistently. Initially uses a C strategy but then switched to a dominant B strategy, however this does not coincide with the child being able to give five items. Accuracy improves or remains perfect for all numbers across sessions (excluding session four).

ID6: Is able to give five items at session five only, this reduces to four items at subsequent testing sessions. Uses a dominant B strategy. Accuracy is mixed, being consistently poor when asked for six or seven items. When asked for three items accuracy is good, reduces, then improves. Accuracy when asked for four or five items does improve over sessions.

ID7: Is able to give seven items at session two, but subsequently only able to give three items for the remaining sessions. Uses a dominant B strategy. Accuracy improves or remains perfect across the sessions.

ID8: Is able to give five at session three, but subsequently only able to give four items. Initially uses a C strategy but then switched to a dominant B strategy. Accuracy improves for all numbers lower than six. Higher numbers accuracy remains poor.

ID9: Unable to give more than four items during any testing session. Uses a dominant B strategy. Accuracy improves or remains perfect across the sessions for numbers lower than seven.

ID10: Is able to give six items on session one but subsequently is only able to give three or four items. Uses a dominant B strategy. Accuracy gets worse across the testing sessions.

ID11: Is able to give five items on session four. Primarily uses a B strategy (except during session one where a C strategy was used), the B then remains a dominant strategy. Accuracy improves across the testing sessions. However, when asked for six items accuracy remains poor.

ID12: Unable to give more than four items. Uses a dominant A strategy. Accuracy is consistently low when asking for three items and does improve for all other requested numbers across the testing sessions.

ID13: Is able to give five / six consistently after session one. Uses a dominant B strategy. Accuracy improves for all requested numbers across testing sessions.

Analysis of children that progress to five knowers during the testing period

ID1, ID5, ID11 and ID13 all progress to be five or more knowers during the testing sessions. ID1 & ID5 both attend nursery setting S, ID11 attends nursery setting R and ID13 attends nursery setting B.

ID1, ID11 and ID13 primarily use a B strategy throughout the testing sessions, whereas ID5 transitions from using a C strategy to a B strategy.

ID13 is able to give five items during testing session one, and both ID1 and ID5 are able to give five items at the second testing session. However, none of the children are able to consistently give seven items until the end of the testing sessions.

All four children's inaccuracy scores improved as the sessions progressed, even after they were able to give five items.

It is important to note that other children were able to give five items on occasion during their testing sessions. However, their performance / knower-level dropped below five for any subsequent testing sessions.

Discussion Chapter

Main findings

This thesis addressed current issues regarding the development of cardinal proficiency in young children. The first study, a meta-analysis, addressed methodological issues exploring the central test of cardinal proficiency. The consistency of implementation, and the effects of different implementation methods were assessed across the extant literature. The second study examined the effects of implementation experimentally. Finally, the third study addressed theoretical issues directly, and explored patterns of development micro-genetically. Crucially, all three studies were driven by analysis of current theoretical assertions regarding cardinal proficiency development. They consistently pinpointed variability in children's development, between tasks and over time, something that current theoretical positions do not account for.

Experimental chapter one presented a meta-analysis comprising 118 studies that utilised the Give-N task (the main test of cardinal proficiency). Here we answered the research questions: *Are there different ways the Give-N task is implemented? If so, do these differences affect children's displays of cardinal proficiency?* The findings from this study suggest that children's achievement during the Give-N task is highly variable, and the way the task is implemented may account for this variability. These variations in children's displays of proficiency is something current accounts of development do not incorporate. Furthermore, with the ubiquitous nature of this task, permeating all areas of developmental psychology, the differences in task implementation and the effects these have on children's displays of proficiency is an important consideration for the field. Comparisons and associations made between early developmental skills may not be valid.

Experimental chapter two explored the effects of implementation experimentally. Children were presented with three versions of the Give-N task. A traditional Give-N task, where children were asked to give a cuddly toy different numbers of items. A role-play version, where children played the role of a shop keeper and were asked, by the customer, for different numbers of fruit and vegetables. The third version of the task was a naturalistic task, where children were asked to pass different numbers of pegs to help hang up washing. Although the findings from this study did not reach significance i.e., no single task produced superior results, children displays of cardinal proficiency did vary between presentations. Some children performed better on the traditional task, some on the role-play and some on the naturalistic. The effects reported in this study supporting the findings from the meta-analysis.

Experimental chapter three explored cardinal proficiency development micro-genetically, testing children weekly over a three-month period. Within this study we directly examined key tenets of current theoretical accounts of cardinal proficiency development, and answered the research questions: *Do children generalise cardinal proficiency abilities to all numbers in their count list and to other cardinal principle tasks once they are able to give five items? Is there variability in children's displays of cardinal proficiency over time? And subsequently, does the use of alternative measures support a gradual view of cardinal proficiency development?* The findings here suggest that conceptual generalisation of the cardinal principle is not a sudden developmental change and support a recent challenges to current accounts of cardinal proficiency development, suggesting a more gradual view of how this skill develops. Furthermore, we found variability in children's displays of cardinal proficiency over time. Again, these variations in children's displays of proficiency is something current accounts of development do not incorporate.

Overall, the thesis asserts that the evidence provides theoretical challenges to current accounts of cardinal proficiency development. Likewise, attention is drawn to the variable implementation of a key cardinal proficiency measure, the effect this can have on children's displays of cardinal proficiency, and the impact this may have on the developmental field.

Theoretical Contributions

Broadly, this thesis contributes to our understanding of cardinal proficiency development and highlights the complex nature of how this key skill may develop. Current theoretical accounts report a straightforward path to understanding cardinality, with children progressing from one level of understanding to the next, in a linear fashion (Le Corre et al., 2006; Lee & Sarnecka, 2010, 2011; Sarnecka & Carey, 2008; Sarnecka & Lee, 2009; Wynn, 1990). However, within the thesis we have evidenced the variability within which children acquire cardinal proficiency, and how small changes within the same task can influence children's displays of this skill. Likewise, during the period of development or change children's proficiency is variable from week to week. This variability is something current accounts of cardinal proficiency development do not account for or discuss. This matters because current methodology primarily utilises cross-sectional studies and the primary measure, the Give-N task, is implemented in many ways, as revealed in study one. Therefore, results from such research may not be reliable. The ubiquitous use of this methodology permeates the developmental field with theoretical assertions based upon such results (see Thomas et al., study one). Consequently, we highlight the need to re-visit current accounts of cardinal proficiency development and draw attention to these methodological issues when assessing cardinal proficiency in future research.

In addition to the above, evidence from this thesis challenges two key tenets of current theoretical accounts and aligns with emerging evidence of an alternative, gradual, account of development (Krajcsi & Fintor, 2023; Krajcsi & Reynvoet, 2023). This will be discussed in detail within the following sections.

Knower-Levels Account

The dominant, 'knower-levels' account of early number development suggests that children progress through incremental stages during development of the cardinal principle (Lee & Sarnecka, 2010, 2011; Le Corre et al., 2006; Sarnecka & Lee, 2009; Wynn, 1990). When assessed using the primary measure of cardinal proficiency, the Give-N task, initially preschoolers are only able to give one item when asked, and at this stage they are deemed a 'one-knower'. Then a month or so later children progress to giving two items and become a 'two-knower', and so on, until they are able to give five or more items. According to the knower-levels account, at this point children undergo an abrupt conceptual change and now understand the cardinal principle – that the final number of a count represents the set size (Chu et al., 2016; Geary et al., 2018; Lee & Sarnecka, 2010, 2011; Le Corre et al., 2006; Sarnecka & Lee, 2009). Following this conceptual advance they quickly generalise this knowledge and are able to give all numbers of items within their count list, and their cardinal proficiency is equivalent across different cardinal principle tasks (Lee & Sarnecka, 2010, 2011; Le Corre et al., 2006; Sarnecka & Lee, 2009). The two key tenets being once children are able to give five items, they a) generalise this knowledge to the rest of their count list (i.e., during the Give-N task they are able to give all numbers of items they are able to recite), and b) generalise this knowledge to other cardinal principle tasks (i.e., there performance on other cardinal principle tasks mirrors their performance during the Give-N task).

Challenges to Count List Generalisation. Within study three we experimentally assessed children's cardinal principle knowledge micro-genetically over a three-month period. The results from this study showed nearly all children during this time period were able to give five items at one or more timepoint. Importantly, many of the children regressed at subsequent timepoints. For some this was temporary, for others this lasted for a number of weeks, suggesting that apparent regressions are developmental patterns rather than behavioural noise. Likewise, once children had given five items at a particular session, they were not automatically able to give higher numbers at future sessions, despite being able to recite the number sequence up to ten (a pre-requisite to our study). The findings here do not correspond with the knower-levels account of development. It does not follow that once children are able to give five items, they undergo a sudden conceptual change and gain insight into the cardinal principle. It may be that development is more gradual, or conceptual change is individual with different children gaining insight at different stages e.g., some children may gain insight when able to give four, others when able to give seven. Additional findings within study three favour the gradual account and will be discussed in a subsequent section.

Irrespective of our favoured account of development our results do not support a key tenet of the knower-levels theory, or at best our results are not accounted for within the knower-levels theoretical position. As discussed previously, variability in children's displays of cardinal proficiency are not incorporated into the current knower-levels account, and our results clearly show variability over time. Development is not as linear or 'neat' as the knower-levels account posits. It is important that we accommodate this into accounts of cardinal proficiency development, as this may need to be considered when assessing this key skill in future studies.

Challenges to Task Generalisation. In all three studies presented in this thesis we find evidence to challenge the generalisation tenet within the knower-levels account of cardinal proficiency development. Both generalisations tenet are grounded in the belief that cardinal proficiency is a sudden conceptual change which occurs when children are able to give five items (Lee & Sarnecka, 2010, 2011; Le Corre et al., 2006; Sarnecka & Lee, 2009). As such, a child should perform equally on any cardinal principle task. The results presented within this thesis suggest this is not the case. Within study two children completed the Give-

N task and the What's on the Card task over several testing sessions, and in study three children completed the Give-N task, the What's on the Card task and the How Many task on the same day and over several testing sessions. Our results revealed that performance on one task does not predict performance on another task, directly challenging the generalisation tenet within the knower-levels theory. Furthermore, results from study one and two suggest that even different versions of the same task[4], in our case the Give-N task, and presenting versions of the same task in different orders, can produce disparate results. Importantly, no one version of the task improved displays of cardinal proficiency. Children's performance appeared to be influenced by different versions of the task within study two, and different cardinal proficiency tasks within study three, therefore such influences cannot be due to procedural demands of the task, which is an argument posited regarding performance on the How Many task verses the Give-N task (Gelman & Meck, 1983).

Again, our results do not support a central tenet of the knower-levels account of cardinal proficiency development, and again the variability we find within children's displays of cardinal proficiency is something not accounted for by the knower-levels theory. The methodological and theoretical considerations for future studies are, as mentioned above, apparent.

An Alternative Gradual Account. The results presented within this thesis suggest that children's routes into understanding number may be more complex than dominant, current accounts posit. The linear stages discussed in the knower-levels account, with a sudden conceptual insight when children are able to give five items may be limited in its explanation. The high variability in children's displays of cardinal proficiency seen in all three studies presented suggest that children's developing knowledge of the cardinal principle

---

[4] Whilst the results did not reach significance, we did note children performing at different levels across the tasks.

is delicate and sensitive to even small changes in task presentation, even after children are able to give five items. Whilst our evidence, on its own, does not deliver a complete, revised theoretical account, it does correspond with emerging evidence suggesting a more gradual development of the cardinal principle (Krajcsi & Fintor, 2023; Krajcsi & Reynvoet, 2023).

The gradual account of cardinal proficiency development parallels the knower-levels account, acknowledging that cardinal principle development is a conceptual change. However, this account removes the confines of a sudden conceptual development when children are able to give five items. This account discusses six knowers or even seven knowers (CP knowers as per the knower-levels account), whose cardinal principle knowledge is equivalent to that of three or four knowers (non-CP knowers as per the knower-levels account) - i.e., a seven-knower's performance on equivalence tasks is not secure, which would be expected if they were cardinal principle knowers (five or more knowers) as per the knower-levels account. As such, the authors posit a more gradual account of conceptual development, with children developing the cardinal principle at different timepoints in their developmental journey (Krajcsi & Fintor, 2023; Krajcsi & Reynvoet, 2023).

The variability reflected in our results are more naturally integrated with this account, as opposed the somewhat strict, linear stages discussed in the knower-levels account. Our strategy assessment measure in study three suggested there was no sudden change in strategy use when children were able to give five items, or at any point during the testing sessions. Specifically, we saw no abrupt advancement in the ways children approached the Give-N task. This aligns with a more gradual view of cardinal proficiency development and, although not specifically explored within this thesis, it may be that the strategies used by children gradually help them to acquire cardinal proficiency. The main strategies adopted by children in study three focused on counting, the required strategy to pass the Give-N task, and this remained consistent across the testing sessions regardless of their success on the task. With counting a dominant strategy across the sessions and across the nursery settings it is likely that children acquire knowledge of the cardinal principle gradually, through instruction and practice.

In addition, our inaccuracy score measure within study three suggested a gradual honing of accuracy when children provided items during the Give-N task. Using this measure as an alternative to traditional knower-levels has facilitated an alternative view of development, outside of the categorical confines of the knower-levels account. Inaccuracy scores for all numbers requested decreased gradually over time, and this was the case for known or unknown numbers, suggesting that children's knowledge and understanding of number is gradually becoming more secure and therefore more accurate.

Importantly, by diverging from traditional measures of cardinal proficiency we are able to view the patterns of development in new ways and provide support for the emerging gradual view of cardinal proficiency development.

Theoretical Conclusions

Overall, results from this thesis suggest that children's development of the cardinal principle is variable, and their patterns of development are not linear and 'neat' as suggested by the dominant knower-levels account. We find that even small variations in task presentation can influence their displays of cardinal proficiency, and their proficiency can vary from task to task in a non-systematic way. This is something that current accounts of cardinal principle development do not incorporate. Furthermore, these findings challenge the central tenets of the knower-levels account. The above-mentioned issues suggest that the current account of cardinal proficiency development requires modification, and perhaps a shift away from the strict confines and the methodologies used to assess development is timely.

Our results begin this shift and correspond with emerging evidence suggesting a more gradual account of cardinal principle development (Krajcsi & Fintor, 2023; Krajcsi & Reynvoet, 2023). Whilst the results from this thesis do not provide a complete and fully evidenced alternative account, they do provide a foundation for further exploration, and alternative measures to assess how this important skill develops.

Methodological Considerations

The thesis comprised a systematic review and meta-analysis, an experimental repeated-measures behavioural study and a micro-genetic behavioural analysis. The three different designs all made use of, or reported on, the existing litmus test for cardinal proficiency, the Give-N task. While two different modifications of this task were utilised in study two the underlying method was preserved to facilitate explication of why differences in presentation may influence children's displays of cardinal proficiency. All three studies and methodologies built upon the previous, triangulating our findings of variability.

Meta-Analytic Design

Study one, adhered to PRISMA and NIRO guidelines for meta-analytic studies, thus improving the rigour and comprehensiveness of this design, and providing an unbiased and objective review of the Give-N task. As with all meta-analyses the comparison of studies is limited, as potential unknown inconsistencies between research groups or between studies cannot be fully accounted for. Nevertheless, this provided a secure grounding from which study two was constructed.

Repeated-Measures Design

Studies two and three adopted repeated-measures designs and whilst this was chosen to reduce between-subject variability, there are limitations that arise from this design. Repeated testing can induce the 'screw-you' effect, where children may tire of completing the same task multiple times and perform sub-optimally due to frustration. Whilst this is possible our observations suggest this is not the case. We ensured one week between each testing session to reduce repetitiveness, and stickers were provided for participation. Verbal feedback from nursery staff suggested that children were always happy and excited to participate, thus indicating a lack of boredom or frustration and a low chance of the 'screw-you' effect. Conversely, repeated testing may prompt improved performance over the sessions through learning. Again, whilst we cannot be sure this was present within our studies, this would be unlikely as children were given no feedback regarding the correctness of their response.

Finally, we ensured the same experimenter undertook every testing session for both studies. This was to reduce any experimenter confounds. However, it is possible that children formed a bond with the experimenter, particularly in study three as testing sessions were over a three-month period. Children's performance may have been enhanced over time as their familiarity with the experimenter increased. Again, whilst this is a possibility, our results do not reflect this. We see variability across testing sessions, so any effects of the experimenter appear low.

Micro-Genetic Design

Other than the repeated-measures considerations mentioned above, the micro-genetic design was able to provide some valuable, detailed and rich evidence tracking the developmental patterns of cardinal proficiency. However, the time-period of this design was limited. By assessing children's knower-levels prior to participation we hoped that including only three or four-knowers within our micro-genetic analysis would ensure the sufficiency of a three-month testing period. Unfortunately, this was not the case with a large proportion of the children not progressing above the four-knower level. Furthermore, returning to the variability we find within children's performance on cross-sectional studies the initial selection of participants may have been unreliable. Nevertheless, adopting the alternative measures (strategy use and inaccuracy scores) during the analysis did allow different assessment of these ‘non-progressors’. Furthermore, their data still showed variations, and at some timepoints they were able to show signs of cardinal principle development by giving five items. Data such as these are still valuable within our micro-genetic analysis.

Outcome Measures

Within the two experimental studies (two and three within this thesis) knower-levels were used as a primary measure of cardinal proficiency, as is standard through the literature. However, within the third study we adopted two alternative measures to assess cardinal proficiency development, these were strategy use and inaccuracy scores.

Strategy Assessment. Assessing the strategies that children use when undertaking tasks is not characteristic of studies within cardinal proficiency research. However, when strategies have been assessed previously, a dichotomy of grabbing verses counting has been reported, with counting classified as the superior strategy and only used when children understanding the cardinal principle. Prior to such understanding the assumption is that children grab numbers of items (Chetland & Fluck, 2007). Given the lack of research assessing strategy use, and the restriction of this study to analyses of dichotomous categories, we decided to examine strategy use in more depth. Guided by a dynamic systems approach and adopting micro-genetic analysis (van Dijk & van Geert, 2007) we felt that a broader approach to strategy assessment may offer an alternative measure of children’s routes to understanding number. We hoped that this would facilitate a deeper understanding of the pattern of developmental change. Furthermore, if children gain cardinal principle insight when they are able to give five items, as per the knower-levels account, then a change in strategy use should coincide.

Strategy assessment, however, did not provide the expected results. Whilst strategy analysis did allow us to test a central tenet of the knower-levels theory, on its own it did not provide the insight expected regarding patterns of developmental change. Strategy use was relatively stable for all children, across all testing sessions. Therefore, our analysis of how cardinal principle knowledge develops was limited. Nevertheless, this measure has provided some ideas for future directions of this research. We suggest that perhaps children are provided with strategies for counting from caregivers and educators, and it is when they utilise such strategies and begin to succeed that knowledge is gradually acquired. However, these assertions need further exploration. In sum, although the strategy measure may not have been fruitful in one respect, it may prove valuable in future studies when explicating the relationship between instruction and development.

Inaccuracy Scores. These were devised as an alternative to knower-levels and are calculated as the difference between the number requested and the number given by children during the Give-N task. Inaccuracy scores were developed to provide a more sensitive measure of cardinal proficiency development, as opposed the more coarse, categorical judgements provided by knower-levels. By assessing the inaccuracy scores per trial this preserves the rich performance detail, as opposed just providing a knower-level for each session. For example, if we consider the knower-levels criterion for success – a child needs to succeed on two out of three trials for a given number for them to be credited with knowledge of that number - we may have two children, both of whom were classified as seven-knowers. However, one progresses through numbers two to seven with no errors, whereas the second child progresses through numbers two to seven making a number of errors along the way and requiring three trials to be administered each time. These children are distinctly different in their routes to success, suggesting that the security of their knowledge of the cardinal principle may also be different. This disparity and detail would be lost within the knower-levels measure. As such, the inaccuracy measure should allow a more accurate measure of cardinal proficiency and facilitate a deeper insight into children’s patterns of development.

Within study three this measure appears less variable than knower-levels measure, suggesting this measure is more accurate, and this measure also supports a more gradual view of development. However, inaccuracy scores, to our knowledge, have not been used to measure cardinal proficiency before and as such have not been rigorously assessed. Nevertheless, we advocate the use of inaccuracy scores for a more accurate view of development, and future studies can then assess reliability of this potentially insightful measure.

Future Directions.

The three studies within this thesis have demonstrated that current theoretical accounts of cardinal proficiency development may need to be re-visited. We have provided support for a more gradual, variable view of cardinal proficiency development, suggested an alternative measure to assess cardinal principle development and posited that instruction, strategy, and development may be linked. In response to our findings future studies could assess the robustness of the inaccuracy measure and focus on explicating the relationship between the strategies children use through instruction and the effects this may have on development of the cardinal principle. A micro-genetic training study could be beneficial to explore these effects.

In addition, reflecting on the third (micro-genetic) study within this thesis, future work could extend our findings by assessing cardinal proficiency over a longer period of time. Despite our selection of only three or four-knowers development over a three-month period saw limited development in some children. Likewise, returning to the variability we find within children's performance on cross-sectional studies the initial selection of participants may benefit from a longer testing period too. Furthermore, future studies should focus on assessing the pattern of development and examining if this is indeed gradual. In addition to a longer micro-genetic study, more participants are needed to allow for growth mixture modelling as this may facilitate a enhanced, evidence based view of the patterns of development.

Finally, it is essential that the variability within children's knower-levels on the Give-N task is considered within future studies. Researchers across the developmental field must acknowledge the variability and account for this when undertaking cross-sectional studies.

Conclusion

All three studies presented within this thesis have shown that variability is present when assessing children's development of the cardinal principle. Likewise, we have shown that children's emerging knowledge is sensitive, even to small changes in task presentation, and this is not always associated with procedural demands of the task. We have also suggested that cardinal proficiency may be a more gradual developmental pattern than current theory allows, and inaccuracy scores may provide a more detailed measure when assessing this important skill.

In light of the above, the dominant account of cardinal proficiency development needs to be re-visited. A shift away from the inter-twined knower-levels measure and account of development is needed, we need to focus on a gradual pattern of development and allow for variability within children's performance. Likewise, new measures such as inaccuracy scores need to be adopted and tested to facilitate a more accurate and detailed account of development, and most importantly, in light of the variability found, researchers need to acknowledge the limitations of cross-sectional studies when assessing development of the cardinal principle.

References

Adolph, K. E., & Robinson, S. R. (2008). In defence fof change processes. *Child Development, 79*(6). 1648-1653.

Aunio, P., & Niemivirta, M. (2010). Prediciting children's mathematcial performance in grade one by early numeracy. *Learning and Individual Differences, 20,* 427-435.

Baroody, A. J., Lai, M-L., & Mix, K.S. (2017). *Assessing early cardinal-number concepts*. Poster session presented ay the Thirty-ninth Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education, Indianapolis, IN.

Brueggemann, A., & Gable, S. (2018). Preschoolers' selective sustained attention and numeracy skills and knowledge. *Journal of Experimental Psychology, 171,* 138-147.

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

Chernyak, N., Harris, P. L., Cordes, S. (2019). Explaining early moral hypocrisy: Numerical cognition promotes equal sharing behavior in preschool-aged children. *Developmental Science, 22*(1), e12659.

Chernyak, N., Sandham, B., Harris, P. L., & Cordes, S. (2016). Numerical cognition explains age-related changes in third-party fairness. *Developmental Psychology, 52*(10), 1555-1562.

Chetland, E., & Fluck, M. (2007). Children's performance on the 'give-x' task: A microgenetic analysis of 'counting' and 'grabbing' behaviour. *Infant and Child Development, 16*(1), 35-51.

Chu, F. W., vanMarle, K., & Geary, D. C. (2015). Early numerical foundations of young children's mathematical development. *Journal of Experimental Child Psychology, 132*, 205-212.

Chu, F. W., vanMarle, K., & Geary, D. C. (2016). Predicting Children's Reading and

Mathematics Achievement from Early Quantitative Knowledge and Domain-General Cognitive Abilities. *Frontiers in Psychology.* https://doi.org/10.3389/fpsyg.2016.00775 .

Chu, F. W., vanMarle, K., Rouder, J., & Geary, D. C. (2018). Children's early understanding of number predicts their later problem-solving sophistication in addition. *Journal of Experimental Child Psychology, 169,* 73-92.

Clayton S., Gilmore C., & Inglis M. (2015). Dot comparison stimuli are not all alike: The effect of different visual controls on ANS measurement. *Acta Psychologica*, *161*, 177–184.

Clements, D. H., & Sarama, J. (2014). *Learning and Teaching Early Math: The Learning Trajectories Approach* (2nd ed.). Routledge.

Coles, A. T. (2014). *Ordinality, neuro-science and the early learning of number.*. Paper presented at Thirty-eighth Annual Conference of the International Group for the Psychology of Mathematics Education (PME38), Vancouver, Canada.

Crooks, N. M., & Alibali, M.(2014). Defining and measuring conceptual knowledge in mathematics. *Developmental Review, 34*(4), 344-377

Diaz, M. G., & Harris, P. L. (1990). The influence of the imagination on reasoning by young children. *British Journal of Developmental Psychology, 8,* 305-318.

Donaldson, M. (1978). *Children's minds.* Harper Perennial.

Donaldson, M. (1982). Conservation: What is the question? *British Journal of Psycholgy, 73*(2), 199-207.

Geary, D. C., vanMarle, K., Chu, F. W., Rouder, J., Hoard, M. K., & Nugent, L. (2018). Early Conceptual Understanding of Cardinality Predicts Superior School-Entry Number-System Knowledge. *Psychological Science, 29*(2), 191-205.

Gelman, R. (1993). A rational-constructivist account of early learning about numbers and objects. In D. L. Medin (ed.), *The Psychology of learning and motivation,* (pp. 61-96). Academic Press.

Gelman, R., & Meck. E. (1983). Preschoolers' counting: Principles before skill. *Cognition, 13,* 343-359.

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

Ginsburg, H. P. (2006). Mathematical Play and Playful Mathematics: A Guide for Early Education. In D. G. Singer, R. M. Golinkoff, & K. Hirsh-Pasek (Eds.), *Play = learning: How play motivates and enhances children's cognitive and social-emotional growth* (pp. 145–165). Oxford University Press

Gilmore, C., N. Attridge, B. De Smedt & M. Inglis. (2014). Measuring the approximate number system in children: exploring the relationships among different tasks. *Learning and Individual Differences. 29,* 50–58.

Gunderson, E. A., Spaepen, E., & Levine, S. C. (2015). Approximate number word knowledge before the cardinal principle. *Journal of Experimental Child Psychology*, *130*, 35-55.

Jordon, N. C., Kaplan, D., Ramineni, C., & Locuniak, M. N. (2009). Early Math Matters: Kindergarten Number Competence and Later Mathematics Outcomes. *Developmental Psychology, 45*(3). 850-867.

Koponen, T., Salmi, P., Eklund, K., & Aro, T. (2013). Counting and RAN: Predictors of arithmetic calculation and reading fluency. *Journal of Education Psychology, 105*(1), 162-175.

Krajcsi, A., & Fintor, E. (2023). A refined description of initial symbolic number acquisition. *Cognitive Development, 65*, 101288. https://doi.org/10.1016/j.cogdev.2022.101288

Krajcsi, A., & Reynvoet, B. (2023). Miscategorised subset-knowers: Five and six knowers can compare only the number they know. *Developmental Science, 27*(1), e13430.

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 Psychology, 195,* 104846.

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*(2), 130-169.

Lee, M. D., & Sarnecka, B. W. (2010). A model of knower-level behaviour in number concept development. *Cognitive Science, 34,* 51–67.

Muldoon, K., Lewis, C., & Freeman, N. (2009). Why set-comparison is vital in early number learning. *Trends in Cognitive Sciences, 13*(5), 203-208.

Muldoon, K., Lewis, C., & Towse. (2005). Because it's there! Why some children count, rather than infer numerical relationships. *Cognitive Development, 20*, 472-491.

Mussolin, C., Nys, J., Leybaert, J., & Content, A. (2012). Relationships between approximate number system acuity and early symbolic number abilities. *Trends in Neuroscience and Education, 1*(1), 21-31.

Nguyen, T., Watts, T.W., Duncan, G. J., Clements, D. H., Sarama, J. S., Wolfe, C., & Spitler, M. E. (2016). Which preschool mathematical competencies are most predictive of fifth grade achievement. *Early Childhood Research Quarterly, 36,* 550-560.

Nunes Carraher, T., Carraher, D. W., & Schliemann, A, D. (1985). Mathematics in the streets and schools. *British Journal of Developmental Psychology, 3,* 21-29.

Piaget, J. (1952). *The origins of intelligence in children.* New York, Norton & Company.

Posid, T., & Cordes, S. (2018). How high can you count? Probing the limits of children's counting. *Developmental Psychology, 54*(5), 875-889.

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.

Sahin, N., Dixon, J. K., & Schoen, R. C. (2020). Investigating the association between student’s strategy use and mathematics achievement. *School Science and Mathematics, 120*(6). 325-332.

Sarama, J., & Clements, D. H. (2009). *Early childhood mathematics education research: Learning trajectories for young children.* Abingdon, England: Routledge.

Sarnecka, B. W., & Carey, S. (2008). How counting represents number : What children must learn and when they learn it. *Cognition, 108*(3), 662-674.

Sarnecka, B.W., & Gelman, S. A. (2004) Six does not just mean a lot: preschoolers see number words as specific. *Cognition*, *92*(3), 329-52.

Sarnecka, B. W., & Lee, M. D. (2009). Levels of number knowledge during early childhood. *Journal of Experimental Child Psychology, 103*, 325-337.

Siegler, R. S. (1996). *Emerging minds: The process of change in children’s thinking.* Oxford University Press, New York.

Szűcs, D., A. Nobes, A. Devine, et al. (2013). Visual stimulus parameters seriously compromise the measurement of approximate number system acuity and comparative effects between adults and children. *Frontiers in Psychology, 4,* 444.

Thelan, E. (2005). Dynamic Systems Theory and the complexity of change. *Psychoanalytic Dialogues, 15*(2), 255-283.

Thomas, H. L., Lewis, C. & Dunn, K. (in preparation). A Systematic Review and Meta-analysis of the Give-N task as a measure of Cardinal Proficiency.

van Dijk, M., & van Geert, P. (2007). Wobbles, humps and sudden jumps: A case study of continuity, discontinuity and variability in early language development. *Infant and Child Development, 16*, 7-33.

Verdine, B. N., Golinkoff, R. M., Hirsh-Pasek, K., Newcombe, N. S., Filipowicz, A. T., & Chang, A. (2014). Deconstructing building blocks: Preschoolers' spatial assembly performance relates to early mathematical skills. *Child Development, 85*(3), 1062-1076.

Wagner, J. B., & Johnson, S. C. (2011). An Association between understanding cardinality and analog magnitude representations in pre-schoolers. *Cognition, 119*(1), 10 -22.

Weisberg, D. S., & Gopnik, A. 2013. Pretense, counterfactuals and Bayesian causal models: Why what is not real really matters. *Cognitive Science, 37,* 1368-1381.

Wilkey, E. D., & Ansari, D. (2020). Challenging the neurobiological link between number sense and symbolic numerical abilities. *ANNALS of the New York Academy of Sciences, 1464*(1), 76-98.

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

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