Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

Linguistic Influences on Early Numerical Development: Pedagogical and Research Implications from a Scoping Review and Systematic Review

Insa-Annett Freyja Hedemarie Tiaden

A thesis submitted to Trinity College Dublin, the University of Dublin for the degree of Doctor of Philosophy on 13/02/2026

Principal Supervisor: Prof. Kathleen McTiernan
Co Supervisor: Prof. Lorna Carson

School of Linguistic, Speech & Communication Sciences

Declaration

I declare that this thesis has not been submitted as an exercise for a degree at this or any other university and it is entirely my own work.

I agree to deposit this thesis in the University's open access institutional repository or allow the Library to do so on my behalf, subject to Irish Copyright Legislation and Trinity College Library conditions of use and acknowledgement.

I consent to the examiner retaining a copy of the thesis beyond the examining period, should they so wish (EU GDPR May 2018).

Summary

This thesis explores the relationship between number-naming conventions in different languages and number acquisition through a scoping review study, followed by an examination of differentiated instruction as a pedagogical response through a systematic review. The scoping review method was a suitable means of examining a broad and complex topic of research where the corpus of current literature is inconsistent and has not yet been properly synthesised. The systematic review provided the opportunity to examine approaches to mathematics instruction in linguistically diverse educational settings. The first chapter introduces the rationale behind this thesis, then gives a short overview of the relationship between number words and numerical thinking. It then elaborates on the research question, study design, and findings of this work and lastly lays out its structure. The second chapter lays the theoretical groundwork through providing key concepts and theories concerning pre-linguistic numerical thinking. Subitizing – the innate capacity to quickly determine the exact number of up to three or four objects – is covered first, followed by the related idea of cardinality. The ability to discriminate between different numbers of objects and do basic arithmetic is discussed. In order to analyse the many forms of mental processing required for comprehending, interpreting, and working with numbers and number concepts, the chapter makes reference to both ground-breaking and recent studies in these areas.

The third chapter explores the development of the language-related components of numerical capabilities, considering why linguistic differences in number naming and within number naming systems is an important area of research. This chapter defines and describes key aspects that play an important role in the acquisition of

formal numerosity in young children, specifically the decimal system and what makes number words special. Three key concepts of numerical vocabulary within a language are discussed here, namely the inconsistencies of many number naming systems, inversion of number words, and homophony and non-pregenitality. The concepts that will be examined will be illustrated with reference to a range of linguistic examples, followed by a closer look at the examples from Danish and Irish. Chapter 4 shows how the transparency of a number naming system facilitates early learning in young children and how inconsistencies can impede the acquisition of counting words. The cardinality principle and the concept of number-knowers are introduced, which denote the gradual realisation that the final number in a counting sequence represents the total number of items in a set. The significance of semantic singular/plural markers within a language on the understanding of cardinality in young children is discussed next. The concept of dimensionality in relation to numeration systems is presented next, after which the frequently used metaphor of the mental number line is used to explain how numbers are mentally represented. This leads to a look at the development of numerical estimation skills, a thorough examination of representations on the mental number line, and lastly, the effects language has on mental number line representation. Chapter 5 turns to a literature review of differentiated instruction, a pedagogical approach which proposes strategies for providing meaningful classroom support for the acquisition of mathematical skills in contexts of linguistic diversity. Chapter 6 provides the research design of the thesis. Chapter 7 describes the search and selection process for the scoping review, charts the data and provides an analysis of its findings in relation to number-naming, including transparency, inversion, and semantic singular/plural markers influence early numerical acquisition. Chapter 8 reports on the systematic review, search strategy,

and an evaluation of the effectiveness of differentiated instruction as a pedagogical response to the linguistic and cognitive variability. Chapter 9 provides a final discussion and conclusion, exploring the findings from both reviews, detailing the contribution to bridging research gaps and exploring the broader implications for inclusive mathematics education.

Acknowledgements

The path to completing this dissertation has been fraught with both professional challenges and personal adversities. I am profoundly grateful for the enduring support of my supervisors, Prof. Lorna Carson and Prof. Kathleen McTiernan, and my former supervisor Prof. Gessica De Angelis, whose steadfast belief in my capabilities and academic potential encouraged me to persevere.

I must also express my deepest appreciation for my mother, who stepped in with critical financial support, when unexpected personal challenges arose. Her unconditional love and sacrifice over the years have been the foundation of my resilience and success.

I am deeply thankful to my partner for his understanding and for being the sunlight I need.

And last, but not least, I want to remember with gratitude my late dog Karlsson, and my current dog Eska for their invaluable emotional support.

Table of Contents

Declaration *2*

Summary *3*

Acknowledgements *6*

List of Tables 11

List of Figures 11

Chapter 1: Introduction *13*

1.1. Study Context 13

1.2. Rationale 13

1.3. Number Words and Numerical Thinking 16

1.4. Research Questions, Study Design, and Findings 18

1.5. Structure 20

Chapter 2: The Acquisition of Numerical Thinking *22*

2.1. Introduction 22

2.2. Pre-linguistic Numerical Abilities 22

2.3. Subitising 23

2.4. Cardinality 27

2.5. Number Discrimination 30

2.6. Simple Arithmetic 32

2.7. Conclusion 33

Chapter 3: Why Number Words are Special *35*

3.1. Introduction 35

3.2. The Decimal System 35

3.3. Numerical Vocabulary 41

3.3.1. Inconsistencies 45

3.3.2. Inversion 49

3.3.3. Homophony and Non-referentiality 52

3.4. The Example of Danish 53

3.5. The Example of Irish 57

3.6. Conclusion 62

Chapter 4: Acquisition of Number Words and Concepts in Children *63*

4.1. Introduction 63

4.2. Transparency and Learning 63

4.3. The Acquisition of Counting Words by Children 66

4.4. The Cardinality Principle 68

4.5. The Concept of Number-Knowers 70

4.6. Semantic Singular/Plural Markers 75
4.7. Recap of Number Acquisition 76
4.8. Dimensionality 77
4.9. The Mental Number Line 81
4.10. Development of Numerical Estimation 83
4.10.1. Representation on the Mental Number Line 89
4.10.2. Linguistic Effects on Number Line Representation 93
4.11. Conclusion 96
Chapter 5: Differentiated Instruction 97
5.1. Introduction 97
5.2. Differentiated Instruction 97
5.2.1. Rationale for Differentiated Instruction 98
5.2.2. Theoretical Underpinnings 99
5.2.3. Critiques and Challenges 104
5.3. Differentiated Instruction and Mathematical Learning 105
5.4. Differentiated Instruction in Multilingual Classrooms 111
5.5. Differentiation Strategies for Multilingual Classrooms 115
5.6. Conclusion 119
Chapter 6: Research Design 120
6.1. Introduction 120
6.2. Research Design Considerations 120
6.3. Systematic Review Research Family 122
6.3.1. Systematic Reviews 123
6.3.2. Rapid Reviews 124
6.3.3. Meta-Analysis 124
6.3.4. Scoping Reviews 125
6.4. Rationale for Sequential Design 126
6.5. Conclusion 129
Chapter 7: Scoping Review (Phase 1) 130
7.1. Introduction 130
7.2. Arksey and O'Malley Methodological Framework for Scoping Reviews . 130
7.3. Identifying the Research Question 132
7.4. Search Strategy 133
7.5. Information Sources 133
7.6. Search Strings 134
7.7. Inclusion and Exclusion Criteria 136
7.8. Study Selection 136
7.9. Data Extraction 138
7.10. Selected Articles 141

7.11. Charting the Data 115
7.12. Findings 121
7.13. Empirical Foundations 121
7.14. Novel Approaches 123
7.15. Thematic Overview 124
7.15.1. Number Naming Variations 124
7.15.2. Inversion 125
7.15.3. Semantic Singular/Plural Distinction 126
7.15.4. Outside Influences 127
7.16. Consulting with Stakeholders 128
7.17. Discussion of Findings 133
7.18. Conclusion 138
Chapter 8: Systematic Review (Phase 2) 140
8.1. Introduction 140
8.2. Review Protocol 140
8.3. Research Question 141
8.4. Search Strategy 142
8.5. Search Strings 142
8.6. Inclusion and Exclusion Criteria 145
8.7. Study Selection 148
8.8. Characteristics of Included Studies 149
8.9. Findings 156
8.9.1. Mathematical Achievement 156
8.9.2. Adaptive Learning and Personalisation 159
8.9.3. Teacher-Led Differentiation 160
8.9.4. Aptitude: Treatment Interactions and Cognitive Profiles 162
8.9.5. Affective Outcomes 163
8.9. Discussion of Findings 167
8.10. Conclusion 172
Chapter 9: Discussion & Conclusion 173
9.1. Introduction 173
9.2. Restatement of the Research Aims and Question 173
9.3. Combined Pedagogical and Research Implications 178
9.3.1. Pedagogical Implications 179
9.3.2. Research Implications and Future Avenues for Research 183
9.4. Limitations 184
9.5. Concluding Remarks 186
References 189
Appendix 227

Appendix 1: Letter confirming Research Ethics Approval 228
Appendix 2: Scoping Review Summary Table 229
Appendix 3. Analysis of Individual Studies located in the Scoping Review 244
Study #1, Cankaya et al (2014) 244
Study #2, Clayton et al (2020) 246
Study #3, Dowker & Roberts (2015) 249
Study #4: Helmreich et al. (2011) 251
Study #5: Lê & Noël (2020) 253
Study #6A and #6B, Li et al. (2013) 256
Study #7 Magargee & Beauford (2016) 258
Study #8 Villarroel et al. (2011) 260
Appendix 4: Systematic Review Summary Table 262

List of Tables

Table 3.1. Number names between 1 to 100 in Danish, English, Japanese	54
Table 3.2. Irish Cardinal Numbers (1–10)	59
Table 3.3. Collective Numerals for Humans	60
Table 3.4. Mutation Effects Caused by Numerals	61
Table 3.5. Ordinal numbers in Irish	61
Table 4.1 Interaction between mother and child	67
Table 7.1. Inclusion and exclusion criteria	136
Table 7.2. Snapshot of part of the data extraction table	140
Table 7.3. Articles selected for inclusion	142
Table 7.4. Stakeholder interview participants	131
Table 8.1. Inclusion and exclusion criteria	146
Table 8.2. Studies selected for inclusion	148
Table 8.3. Description of methods of selected studies	153
Table 8.4. Selected studies that reported achievement outcomes in mathematics	158
Table 8.5. Effects on attitude outcome	164
Table 9.1. Pedagogical implications for language-informed mathematics instruction	182

List of Figures

Figure 4.1. Different possible number representations	78
Figure 4.2 Predicted estimates of the logarithmic and linear ruler models	86
Figure 7.1. Prisma flow diagram phase 1	137
Figure 7.2. Findings regarding number name transparency	116

Figure 7.3. Languages included in studies 117

Figure 7.4. Occurrence of relevant languages 117

Figure 7.5. Difference in sample sizes across all featured studies 118

Figure 7.6. Areas of thematic overlap within the featured studies 119

Figure 8.1. Prisma flow diagram phase 2 147

Figure 8.2. Countries in which research was conducted 149

Figure 8.3. Methodological designs of the studies 150

Chapter 1: Introduction

1.1. Study Context

Early numerical abilities include basic skills like counting, number recognition, and simple arithmetic. These abilities are foundational for children's skills and development, including their future academic performance not only in mathematics, but also in broader cognitive development (such as working memory and problem-solving). Language plays a critical role in how we conceptualise and process information, as well as the linguistic properties of different language varieties. The structures of languages seem to shape early cognitive abilities, including numerical ones. Number words or number names are an essential part of early number understanding, and it is not simply that children learn to distinguish between words such as "one", "two", and "three". They also are learning to distinguish cognitively between these constructs.

In this thesis, we seek to determine what is special about number names, especially whether and how the linguistic structure of number words influences how children develop numerical abilities. As some languages have a more transparent number naming system, without irregularities, we can compare how these linguistic differences may influence the way children learn to count, understand number relations, or develop concepts of magnitude.

1.2. Rationale

Numeracy and mathematical abilities are extremely important skills to have in our modern societies. Unfortunately, mathematics and arithmetic often seem to

be seen as impenetrable and difficult to learn beyond a relatively basic understanding. One reason for this can be, that learners fail to grasp the full structure of the base-10 system almost universally used today where numbers are concerned. Several global comparisons have shown a tendency for learners in Asian cultures to have an easier time acquiring mathematical skills. One reason that has been proposed to explain that apparent advantage is that many Asian number naming systems are extremely transparent and congruent with the written structure of Arabic numerals. This makes it easy for children to learn and grasp the base-10 system without having to put additional energy into processing or having to learn inconsistent number names by heart.

To find out more about how children's acquisition of numeracy is affected by the particular quirks of their number naming systems is important because, in a globalised world with increased migration and more and more diverse language landscapes in classrooms, potential educational advantages and disadvantages should be identifiable and equalised to create a playing field as equal as possible for all children. As we will discuss later in detail in this thesis, early delays in numerical learning caused by lack of transparency in the number naming system can potentially have a ripple effect that also caused later difficulty learning and using arithmetic and mathematical understanding.

To understand better how phenomena like inversion, number names based on archaic system that differ from the usual base-10 system we use today, and even semantic distinction of singular and plural in a language can have an effect on the initial numerical understanding of young children, means to potentially make mathematical education more accessible for these children later. Being a native speaker of German myself, a language with an almost full inverted number

naming system, and from talking to people who grew up with number names that were in-transparent in other ways, it seems to be the case that native speakers are often largely unaware of the inconsistencies in their number names. This would also mean that maths teachers are largely unaware of these inconsistencies and cannot counteract potential negative effects on children's learning.

When it comes to teaching strategies that can lessen these language influences, some students may have different conceptual starting places when they enter maths classes, influenced for instance by number-naming conventions and cognitive accessibility. A crucial pedagogical concern is raised: how can education be adapted to account for such variation in students' mathematical preparedness?

Differentiated instruction (DI) is one approach to addressing learner variability in the classroom, but there are gaps in research coverage of DI in contexts of mathematics education, especially a lack of studies which compare the use of differentiated instruction to non-differentiated instruction. Therefore, in order to determine the nature of the empirical evidence regarding the effectiveness of DI in mathematics classrooms, a systematic review was undertaken as the second phases of this thesis, focusing specifically on achievement and attitudes in primary and middle school students in mathematics classrooms. It is clear that more research in this thus far largely unexplored area is necessary and this thesis is intended as an informed starting point to facilitate this.

1.3. Number Words and Numerical Thinking

Numerical thinking is an innate ability that exists in animals and new-born babies, but which only allows for the use of a limited and finite amount of the otherwise infinite extent of numbers in existence. To access infinity, human children must learn to count, and they usually do so in their native language(s). However, acquiring number words is distinct from understanding the concept of numbers. Understanding the concept of numbers, in turn, is the basis for mathematical learning and future abilities and can be hindered and delayed by challenges arising during the acquisition of number words. These challenges might plausibly increase for bilingual and multilingual children, but it is as yet unknown whether this is the case and how such challenges might manifest.

Number words can differ greatly from language to language, and a small but growing number of researchers take an interest in whether these differences matter when it comes to acquisition of numeracy. These are for instance: Cankaya et al. (2014), Clayton et al. (2020), Dowker and Roberts (2015), Helmreich et al. (2011), Lê and Noël (2020), Li et al. (2013), Magargee and Beauford (2016), and Villarroel et al. (2011). Together this body of research is directly relevant to the investigation conducted in this thesis. Additionally, further studies also relevant to the study of the interrelatedness of number-naming conventions and the acquisition of numeracy include Butterworth and Gelman (2005), Göbel et al. (2011), Miller et al. (1995), Musolino (2004), and Shaul and Bar (2021). We will return to the work of these researchers later in this thesis. Some of the key questions asked by scholars active in this research field include:

- Do more complex number words slow down the acquisition of number concepts themselves?
- What happens if several languages, and their number words, influence each other in the minds of young children who are in the midst of learning numbers and how to use them?
- To what extent can differentiated instructional strategies mitigate potential challenges arising from differences in number-word structures across languages?
- How do teachers adapt mathematical instruction to account for linguistic diversity, and what impact do these adaptations have on learners' numerical development?

In light of this growing research activity, this thesis examines the effect of number naming conventions in different languages on children's first acquisition of numerical concepts. This is a field of study that has thus far seen very little attention. The original plan for this project was to undertake an experimental study involving a series of tests on location at primary schools in Germany, Italy, and Denmark with bi- and trilingual children. I originally planned to conduct an empirical study on the subject of learning to count in multiple different number naming systems at once and how these different systems might influence each other. Unfortunately, the Covid-19 pandemic and its after-effects made it impossible to find enough schools able to let me visit to conduct my tests and therefore I could not find enough participants. Being confronted with the necessity to change course completely, I found myself looking back at the existing research and was struck again by how little this topic has been explored to date, even where monolingual children are concerned. Therefore, I chose the dual research approach of conducting a scoping review followed by a systematic review to determine what is already known and where there are still gaps in research as a means of contributing to the development of this field. The results and subsequent the closing of these gaps and expansion of research will be of

interest to the fields of psycholinguistics, developmental linguistics, and applied linguistics alike.

1.4. Research Questions, Study Design, and Findings

A sequential research design was employed, to provide for both breadth and depth in a two-part review of current empirical research on number names. Firstly, a scoping study addresses a research gap in the field concerning the effect of number naming conventions and inconsistencies in different languages on the acquisition of early numerical skills in young children. The research question that was addressed in phase one of this research study, the scoping review, is as follows:

What does the literature indicate about the relationship between number-naming conventions in different languages and number acquisition in young children?

The scoping review utilised the review framework as outlined by Arksey and O'Malley (2005), along with the PRISMA-ScR checklist (Tricco et al., 2018), to ensure that article selection decision-making and reporting were carried out transparently and to the highest standards.

The findings of the scoping review demonstrated that in each of the studies selected for inclusion in the scoping review, a connection was found between numerical concepts and the variations in number naming practices and the ease and speed at which young children learn numbers. According to research on number naming transparency, children who have languages with (more) transparent number names seem to learn to count, acquire early numeracy, and

understand place value more quickly (Magargee & Beauford, 2016; Thanheiser & Melhuish, 2019; Guerrero et al., 2020). These children also make fewer inversion errors and have more accurate estimation skills on a number line at a younger age. Children that have a semantic singular/plural distinction in their language are better able to understand cardinality, particularly the distinction between sets of "one" and sets of the following small numbers. A common theme among the selected studies in the scoping review, that took home numeracy and societal factors into account, appeared to be the idea that these external factors towards early learning are just as significant as or even more significant than linguistic characteristics in relation to early numerical acquisition.

A second phase of the research, a systematic review, was conducted in order to address a research gap that was identified in phase one related to learner variability and the comparative effectiveness of differentiated instruction (DI) versus non-differentiated instruction. The research question that was addressed in phase two, the systematic review, is as follows:

What does the literature say about the effect of differentiated instruction (DI) compared to non-DI on mathematics achievement levels and attitudes toward mathematics in primary and middle school students?

Twelve empirical research from four continents were evaluated through a systematic review that adhered to a strict methodological standard. The results showed that Differentiated Instruction efficiently bridges the conceptual gaps caused by the linguistic and cognitive elements discussed in the previous chapter, consistently producing beneficial impacts on both mathematics

achievement and student attitudes. Effective outcomes depend on precise diagnosis, often made possible by adaptive technologies or learning analytics, combined with instructional scaffolding tailored to the learner's cognitive profile, is a recurring topic in the data. Additionally, the review demonstrated that organised, teacher-directed instruction, rather than more autonomy, is frequently necessary for effective differentiation for lower-achieving children. Although teacher-led approaches were shown to be quite successful in enabling curriculum access in a variety of contexts, the synthesis also pointed out a sustainability issue, pointing out that without ongoing professional assistance or technological integration, achievement gains frequently fade.

1.5. Structure

When children acquire language, they learn different linguistic numerical concepts through different acquisition processes and in predictable stages they form mental number representations. These can be affected by inconsistencies in number naming conventions and even the form semantic singular/plural distinction is realised in a language. The first two chapters in this thesis lay the groundwork for the scoping review by providing key definitions and review of seminal literature related to numerical thinking and the acquisition of number words. These two chapters introduce the critical concepts that form the study's design and its outcomes. Pre-linguistic numerical abilities are outlined first as these abilities are thought to be innate and have been shown to exist in some animals. Pre-linguistic numerical abilities are the basic building blocks of all numerical abilities that will be acquired later by children, without them numeracy is impossible. An overview of what makes number words a category of their own

in terms of language and how they are mentally processed is then outlined. The number naming system of two languages, Danish and Irish, are discussed in greater depth since both are highly opaque, and as a result of that opacity, they illustrate several inconsistencies that also exist in one way or another in many other languages. The following four main concepts are discussed in the literature review: transparency, inversion, outside influences, and semantic singular/plural distinction. The research design chapter presents the rationale for a scoping review followed by a systematic review. The eight articles selected for inclusion in the scoping review are analysed and discussed and in the context of these four main concepts.

Additionally, the educational framework of differentiated instruction (DI) is discussed in the context of mathematics instruction and multilingual classrooms. The twelve articles selected for the systematic review are analysed and discussed from the viewpoint of mathematical instruction. Finally, a concluding chapter synthesises the results from the scoping and systematic reviews, discusses these findings, and examines the pedagogical and research implications as well as the study's limitations and avenues for future research.

Chapter 2: The Acquisition of Numerical Thinking

2.1. Introduction

This chapter outlines the literature related to the pre-language abilities involved in the acquisition of numerical thinking, starting with subitising, the inherent ability to tell the precise number of up to three or four objects at a glance, followed by the associated concept of cardinality. It then goes on to example the ability to discriminate between numbers of objects and to conduct simple arithmetic. The chapter examines the literature related to the types of mental processing involved in understanding, interpreting and manipulating both numbers and number concepts.

2.2. Pre-linguistic Numerical Abilities

Numerical thinking involves complex cognitive processes. Pre-linguistic numerical abilities are a fundamental prerequisite for the later acquisition of linguistic numerical abilities. The development of numerical thinking in children begins from an early age (Wiese, 2003), and involves early number recognition, counting and numerical relationships. The following four numerical abilities have been identified by researchers (e.g. Jensen et al. 1950; Koechlin et al., 1997; Mandler & Shebo, 1982; Suzuki and Kobayashi, 2000; Piazza et al., 2002; Wynn et al., 2002; Piazza et al., 2003) in a variety of animals and pre-linguistic children:

- Subitisation
- Cardinality
- Number Discrimination
- Simple Arithmetic

The following four subsections define each of these in turn, concentrating on how these abilities manifest in human infants.

2.3. Subitising

Subitising is one of several basic numerical skills which a range of different animals share with humans, especially human infants. Long before they learn to speak, babies can already subitise, recognise the cardinality in sets to up to three or four items, and also discriminate between the lowest numbers and are even able to do basic arithmetic. This indicates that numerosity is a system distinct from language and in learning to count, several systems need to be combined. It is argued that every human being has the ability to subitise, in other words to tell at a glance whether one, two, or three items are in front of them, regardless of their background culture or their level of education (Piazza et al., 2003).

Without training, however, humans cannot easily identify a large number of objects. Typically, four items are the most we can readily distinguish, with higher numbers requiring more time and cognitive work, such as sorting items into smaller, subitisable groupings and counting those. In 1949, Kaufmann et al. coined the term subitising and he had already observed that it was a process separate from calculating or counting. The terms subitisation and subitising are derived from the Latin word *subitus*, meaning 'sudden' (Jensen et al. 1950; Mandler & Shebo, 1982; Piazza et al., 2002). The term itself is a little misleading, however, because subitisation, no matter how quickly it takes place, is far from instantaneous.

Identifying a group of three dots takes around five or six tenths of a second, or about the time it takes to read a word aloud or recognise the face of a familiar

person. This duration is also not constant: it gradually grows from number 1 to number 3. Hence, subitisation probably requires a series of visual operations, all the more complex the greater the number to be recognized (Dehaene, 2011). Still, for all intents and purposes, subitising is a process that takes place instantaneously, automatically, reliably, and without conscious attention. Objects are not the only things we can subitise – we also do it with flashes of light or sounds like beeps or knocks (Davis & Pérusse, 1988).

Subitising involves different cognitive processes than counting. We recite sequences of words while counting, whether aloud or only in our minds, and so our ability to count, but not our ability to subitise, should be affected when asked to do an articulatory task and a cardinality assessment at the same time. Hitch et al. (1987) confirmed this when they showed pictures of multiple houses on a computer screen to children. The children were instructed to click the button on a numerical keyboard that indicated the number of houses they saw, as accurately and as quickly as possible. Additionally, half the children were also told to say "blah blah blah blah" repeatedly at the same time. The extra articulatory task had a profound impact on speed and precision, as these children were slightly slower and made more mistakes than the control group who performed the task in silence. The disparity, however, existed only for quantities greater than three, with no disruptive effect apparent for sets with three or fewer houses. This outcome suggests that counting does include subvocalising, while subitising does not.

It is also clear from the amount of time needed for cardinality assessments that subitising and counting are distinct procedures. Subitising happens almost instantaneously yet very accurately, but it works only for sets of up to three, and

at most four, objects. The time it takes to determine the cardinality of sets containing one, two, or three items is roughly 40 milliseconds per item, which is presumably the time it takes to individualise each item during the process. Reaction times increase drastically to around 300ms per object for sets of more than three or four, likely the time the cognitive counting process takes to assess (Wiese, 2003). This phenomenon, the swift, almost immediate, evidently effortless, but still precise assessment of cardinalities of small sets of up to three, that does not require counting procedures, was dubbed the 'discontinuity after three' by Fischer (1992). For him, it was evidence of a specific apprehension mechanism pertaining to small cardinalities.

It is not possible to easily and clearly distinguish cardinalities above three or four when there is no counting involved as they become fuzzy and their representation seems ongoing. This means that the sizes of bigger sets are so-called ‘noisy magnitudes’, difficult to determine accurately due to uncertainty. In magnitude estimations of larger sets, response time and standard variation increase in direct proportion to the size of the set (Whalen et al., 1999). Additionally, this means the bigger the sets, the more difficult it becomes to determine whether their sizes are too close together (e.g. seven or nine items). The discrepancy in number of objects in large sets must be sufficiently largeto be discernible, and the larger the sets, the greater this discrepancy must be. When comparing their cardinalities, the closer the sets are in number and the greater their sizes, the less accurate the estimation will be.

Sathian et al. (1999) found that, at least in part, the brain regions involved in subitising and counting are disjointed, indicating that a difference in neural underpinnings reflects the subitising and counting distinctions observed. His

participants were seated in front of a computer screen on which between one and eight vertical bars were displayed. Their job was to say how many bars they could see each time, as quickly as they could. Sathian and his colleagues used position emission tomography (PET) to measure the regional cerebral blood flow of the participants as they were performing the task. Increased cortical blood flow suggests elevated activity in the brain regions where it occurs. While comparing PET images of cardinality assessment, they discovered that regions that were active during assessments from five to eight items were not active during assessments from one to three items and vice versa. It seemed to be down to individual differences in which category the assessment of four items fell. These findings suggest that, on the neurological level, the distinct nature of mental processes involved in subitising or counting visual objects is reflected to some degree.

There is a strong agreement that subitising is not just a method of pattern recognition. Researchers have now been examining for more than a century which neural system is implicated, and especially whether small numerical quantities are represented by a unique subitising system that is different from a large number estimating system. Gelman and Gallistel (1986) believed subitising to be serial processing with visual attention to each particular object, or in other words, counting very quickly. On the other hand, Dehaene (2011) proposed that it is conducted by parallel pre-attentive processing, which does not require individual attention to each item. He cites evidence from people with brain injuries that prevent them from serially paying attention to items around them, and thus counting them, but who can nevertheless easily and reliably subitise up to three objects (Dehaene & Cohen, 1994). Recently Kuttner et al. (2023)

demonstrated that the representation of small and large numbers is driven by two distinct brain pathways. They recorded single neurones for the purpose of number judgement in the medial temporal lobe of neurosurgical patients. When counting reached four, they discovered a border in neural coding that corresponds with the behavioural shift from subitising to estimation. Neurones exhibited enhanced tuning selectivity in the subitising range, along with suppression effects that may indicate surround inhibition as a mechanism for enhancing selectivity. On the other hand, a ratio-dependent number estimation system was characterised by a decline in tuning selectivity as the numbers increased above four. Using decoding and clustering analysis, they also identified the coding barrier dividing the two systems and concluded that the small number subitising system may be related to working memory and attention, since it exhibits similar capacity restrictions.

Subitising and cardinality are closely associated. As we have seen, subitising describes the ability to judge the number of items in a small set without the need for counting. The following subsection outlines the literature related to cardinality, which is a fundamental concept in mathematics that allows us to quantify the size of a set.

2.4. Cardinality

In 1961, Pastore taught canary birds to choose the fourth or fifth tablet from a sequence of tablets of aspirin. The pills were randomly and unevenly dispersed over many cubicles along a runway, so that the overall number of pills the canaries had experienced so far was the only indication they could use. In an expanded paradigm during an experiment by Davis and Bradford (1986), which

was later repeated by Suzuki and Kobayashi (2000), rats were taught to always enter a tunnel with a particular number in a labyrinth, e.g. always the fifth tunnel or a box of a specific rank in an array of boxes. The set-up was designed in such a way as to make it difficult for the rats to use positioning as clues. For example, to deter them from using the absolute position of the tunnel within the labyrinth, the lengths between tunnels would vary and the total number of tunnels would change to prevent them from using the relative location of the tunnel within the sequence.

The ability of the rats and birds to perform these tasks successfully implies that they were able to develop a concept of sequential order independent of physical location. This is the same concept that human beings refer to in ordinal number assignment, and which is involved when numbers in a numerical subroutine, such as counting, are used. This is a basis for cardinal number assignment (Wiese, 2003). All of this suggests that our concept of cardinality can build on an evolutionary older capacity that humans share with other species, in particular with other warm-blooded vertebrates.

Cardinality is not a characteristic of objects, as colour or consistency, but rather a property of sets. Sets are made up of groupings of items that are captured under a certain notion. To put it another way, cardinality is based on categorising items according to a unifying concept that applies to every item in the set. Take a box holding six toys, for example: three dolls, two toy cars, and a ball. Cardinality is determined by how the mind groups these items. For example, if the relevant concept is toys, the set has a cardinality of six, but if the relevant concept is types of toys, the set has a cardinality of three, and so on. As Frege spelled out in a crucial paragraph of his *Foundations of Arithmetic* in 1884, this difference holds

true not only for cardinalities, but for every number attribution:

It is quite true that, while I am not in a position, simply by thinking of it differently, to alter the colour or hardness of a thing in the slightest, I am able to think of the Iliad as one poem, or as 24 Books, or as some large Number of verses. Is it not in totally different senses that we speak of a tree as having 1000 leaves and again as having green leaves? The green colour we ascribe to each single leaf, but not the number 1000. If we call all the leaves of a tree taken together its foliage, then the foliage too is green, but it is not 1000. To what then does the property 1000 really belong? It almost looks as though it belongs neither to any single one of the leaves nor to the totality of them all; is it possible that it does not really belong to things in the external world at all? (Frege, 1884, p. 522)

Subitising suggests that some notion of cardinality is independent of language. Subitising allows for almost instantaneous, accurate, and specific representation for small sets of one, two, three, occasionally four, objects. However, we rely on rather fuzzy, continuous, and uncertain representation for sets with numerical quantities greater than three or four. It is therefore counting that imposes a completely different idea of cardinality, regardless of the size of a given set, and it is counting which essentially enables one to generalise the notion of correct quantification to sets of any size.

To be able to distinguish between the cardinalities of small sets, i.e. subitise, or even to work out that three minus two equals one, language is not required. The capacity to figure out the relationship between small amounts without using numerical tools, which includes counting, has been found in human babies, as well as some animals such as non-human primates, raccoons, rodents, dolphins,

parrots, crows and pigeons (e.g. Koehler & Wachholtz, 1936, 1950; Merchner, 1958; Merchner & Guevrekian, 1962; Rumbaugh et al., 1987; Matsuzawa, 1997; Suzuki & Kobayashi, 2000). Piaget et al. (1941) claimed that children under the age of four or five had only little comprehension of the cardinality of sets, much research has been completed since then with pre-schoolers, babies, and even new-borns, resulting in a great amount of evidence that subitising is an intrinsic skill (see Mandler & Shebo, 1982, for a review of subitising experiments). The following subsection outlines the experimental evidence from studies that focused on number discrimination in preverbal infants.

2.5. Number Discrimination

This subsection examines some key studies which indicate that neonates and babies have some capacity to discern small numbers, and hence some innate numerical aptitude. Starkey and Cooper (1980) showed slides to newborns aged 4 to 5 months, seated on their mothers' laps. They kept track of how long each infant spent staring at each slide before losing interest and looking away, at which point a new slide was introduced. At first, the slides had two big black spots on them. During the trials, the new-borns were always presented with two dots next to each other horizontally, although at varied distances. Sometime later, they would habituate, i.e. begin to pay attention to the slides for increasingly shorter amounts of time. When the slides were then suddenly changed to ones with three black dots, the new-borns instantly displayed a longer fixation period, i.e. they stared at them for longer again. The experiment was subsequently done in reverse, using three dots first, and then two, with the same result. This shows that babies can detect the difference between two and three dots based on the consistent change in fixation time. In 1983, Antell and Keating achieved the

same outcome with babies as young as three to four days old. The findings were also confirmed using slides of objects of various sizes, shapes, and alignments (Strauss & Curtis, 1981). Infants may even differentiate three to four objects under specific circumstances (Strauss & Curtis 1981; van Loosbroek & Smitman, 1990). How abstract this ability is, is demonstrated by the fact that it is not limited to visuals, nor is it bound to a certain arrangement of items or to the objects' identity. In other words, objects inside a set can be replaced as long as the cardinality remains the same. Starkey et al. (1990) demonstrated that children aged 6 to 8 months can distinguish between sets of pictures displaying diverse items of varied sizes and shapes. Van Loosbroek and Smitman (1990) successfully used a video screen to show babies aged 5 months, 8 months, and 13 months groups of moving images in patterns that grew and shrank in size but whose number remained constant. Six-month-old new-borns reliably differentiated between two and three jumps of a doll in Wynn's (1996) research. By the age of seven months, infants recognise numerical equivalence between a number of items and drumbeats of the same number (Starkey et al., 1990). Even babies as young as three to four days old can distinguish between small numerical quantities, as they can tell apart spoken phrases with two and with three syllables (Bijeljac-Babic et al., 1991).

In young babies, early salient concepts of discrete sounds and physical objects have been demonstrated to be available for set formation. Wynn et al. (2002) showed that babies as young as five months old may very well be able to use the more complex notion of groups. During their experiment, the infants were shown two presentations: one with two groups of moving dots on a computer screen, and another with four groups of moving dots. The babies were able to discern

between these presentations, i.e., sets that comprised groups of items as elements, as opposed to individual objects. The literature related to the fourth pre-linguistic numerical ability, the ability to conduct basic arithmetic operations is presented in the following section.

2.6. Simple Arithmetic

The habituation technique has also been used to establish basic arithmetic abilities in babies (Wynn, 1992, 1998). Four-and-a-half-month-old infants were evaluated using a procedure known as violation-of-expectation-paradigm in developmental psychology. It was discovered that the babies were aware that one plus one is two and that two minus one is one. The methodology was as follows: new-borns were shown one puppet on a stage, then the stage was covered by a screen in front. While the infants were looking at the closed screen, a second identical puppet was placed on the stage behind the screen and the screen then lowered. Now, either two puppets (expected outcome) or just one puppet (unexpected outcome) could be seen on the stage. Babies stared at the stage for considerably longer if there was just one puppet than if there were two puppets, implying that the infants were surprised by something obviously impossible occurring. Similarly, babies gazed longer at a stage when a third puppet seemed to have appeared out of nowhere (Wynn, 1992). The conclusion drawn from this research is that new-borns understand that one plus one equals two, not one nor three.

In a second experiment, infants were first shown two puppets on stage. Then, while the screen was closed, one puppet was taken away. If just one puppet was on stage after the screen was lowered again, the babies were unsurprised, and they did not gaze at the stage for any longer than normal. If there were still two

puppets on stage, however, the new-borns gazed at the result for far longer. This indicates they could tell that two minus one equals one (Wynn, 1998). Infants were found to anticipate that two puppets plus one puppet equals three puppets, and that three puppets minus one puppet equals two puppets at six months of age.

To exclude any location-based expectations influencing the results, Koechlin et al. (1997) repeated the tests with puppets placed on moving turntables and came to the same results. This implies that infants employ mechanisms that are more abstract than object location and this was affirmed by additional research. Simon et al. (1995) repeated Wynn's (1992) experiment but included puppets transforming into balls behind the screen. This change was designed to verify that the results had nothing to do with specific objects, but rather simply with abstract numbers. When the screen was lowered and two balls emerged instead of two puppets, four-to five-month-old infants exhibited no surprise and stopped looking at the stage. They did, however, stare for longer and showing surprise if they saw one or three balls when they were expecting two. This suggests that for four-to five-month-old infants, number was more important than object identity. In contrast, older infants during the same study were also astonished to see two balls instead of the expected two puppets.

2.7. Conclusion

This chapter has introduced four pre-linguistic numerical abilities:

- Subitisation: The innate ability to recognise a set of up to three or four items instantaneously, automatically, reliably, and without conscious attention.
- Cardinality: The understanding that items can be grouped based on a changeable unifying concept that applies to every item in the given set.

- Number discrimination: The innate capability to differentiate one small number of items from another small number of items.
- Simple arithmetic: The expectation that one plus one item equals two items, and subsequent surprise when three or zero items are presented instead.

Building on the concepts outlined in this chapter, the next chapter will turn to the acquisition of the *language*-related components of numerical capabilities. It will explore the difficulties of the acquisition of number words in the first language, why number words are a special and distinct part of a given language and what role linguistic transparency of number naming plays in mathematical learning even long after the initial stages of learning to count and to understand numbers.

Chapter 3: Why Number Words are Special

3.1. Introduction

This chapter explores the development of the language-related components of numerical capabilities, considering why linguistic differences in number naming and within number naming systems is an important area of research. Specifically, the chapter defines and describes key aspects that play an important role in the acquisition of formal numerosity in young children, specifically the decimal system and the uniqueness of number words. It is not presumed, however, that all divergences between languages will lead to difficulty in students' comprehension and connecting the number name and numeral. It is impossible to make any universal statement one way or the other, given the small number of studies conducted on this matter. The concepts that will be examined will be illustrated with reference to a range of linguistic examples from the literature.

3.2. The Decimal System

The decimal system, also defined as the base-10 positional numeration system or the base-10 value system, is the world's most frequently used number system. It is a tool of extraordinary efficiency and efficacy (for taxonomies of number systems, see Chrisomalis, 2004; Widom & Schlimm, 2012; Zhang & Norman, 1995). Learning whole number arithmetic in schools across the world involves learning the decimal system and how to compute using it, even if it is incompatible with traditional number language (Sun & Bartolini-Bussi, 2018).

The reason for its widespread adoption lies in the way it represents numbers. Following Howe (2019), below is a brief overview of the decimal system's fundamentals and the five stages of place value. The five-stage system is a developmentally adapted elaboration of the New Math's academic theory (Chambris, 2015; Sun et al., 2018). To begin, any base-10 number is inherently a sum which can be typed out and broken down into its components using Arabic numerals as follows:

$$472 = 400 + 70 + 2$$

Each number is comprised of a series of special numbers for which Beckmann (2002) coined the term "place value parts" (PVP). Crucially, a PVP has a multiplicative structure, as the digit for each PVP records the number of copies of a base-10 unit that were used to generate that number, for example:

$$400 = 4 \times 100 \quad 70 = 7 \times 10 \quad 2 = 2 \times 1$$

The base-10 units are 1, 10, and 100, and the 1-digit number multiplying each unit indicates how many copies of that unit were utilised to create the number. The number's symbol, written in Arabic numerals, 472, indicates which unit is multiplied by which digit through the digit's position, i.e. the rightmost digit

indicates the number of 1s required, the next digit to the left indicates the number of 10s required, the digit one further to the left indicates the number of 100s required, and so on. The size of the base-10 unit that corresponds to each digit is therefore disclosed by the digit's location in the base-10 representation.

To make this system work, a zero “0” , a symbol that represents nothing, is required so that, if no copies of a given base-10 unit are necessary to represent a given number, one can still write something in the corresponding place and therefore allowing digits representing larger base-10 units to appear in the correct places, for example:

$$509 = 5 \times 100 \quad 0 \times 10 \quad 9 \times 1$$

The 0 in the 10 place functions as a placeholder for the 10-unit, so that the 5 (in this case) remains in the correct position, denoting the 100-unit.

The multiplicative structure of the PVPs is not exhausted by knowing that each component represents a digit multiplied by a base-10 unit. The basic ten units also have a multiplicative connection with one another. The 10-unit is made up of 10x1, the 100-unit is made up of 10x10, the 1000-unit is made up of 10x100, and the 10,000-unit, is made up of 10x1000, and so on. In other words, the number of zeroes used to represent each unit is equal to the number of times the unit is multiplied by 10. This allows for an even more compact representation of bigger base-10 units, by simply noting the number of factors of ten used to create

a particular unit. In algebra, for example, it is commonplace to write:

$$10 = 10^1$$

$$100 = 10 \times 10 = 10^2$$

$$1000 = 10 \times 100 = 10 \times 10 \times 10 = 10^3$$

The number of factors of ten needed to produce one of these products is called the power of ten, also known as the *order of magnitude* or simply *magnitude*. It also denotes how big a particular unit is, with each being 10 times as big as the next smaller one and ⅒ times as big as the next bigger one. As a result, a PVP's order of magnitude simply refers to the magnitude of its base-10 unit, which is equal to the number of zeroes in the part's standard representation. For instance, 2 has magnitude zero, 70 has magnitude one, 400 has magnitude two, and so on. To summarise, the conventions of the decimal system contain the following:

- Each number is composed of place values parts (PVPs)
- Each of these components is a single-digit number multiplied by a base-10 unit
- The location of a particular number indicates with which base-10 unit it is multiplying
- Furthermore, each base unit in itself embodies a multiplicative property, namely it is the forming power of the system's base, which in this case is ten.

In order to make the complete decimal structure explicit, it must be broken down into these five stages during learning, based on the powers of 10 for each digit's place value:

1. 472

2. $= 400 + 70 + 2$

3. $= 4 \times 100 + 7 \times 10 + 2 \times 1$

4. $= 4 \times (10 \times 10) + 7 \times 10 + 2 \times 1$

5. $= 4 \times 10^2 + 7 \times 10^1 + 2 \times 10^0$

According to Howe (2019), it is critical to recognise that each stage reflects a conceptual advancement over the preceding one, and that each stage may take one or more years to cognitively develop. The standard way of writing numbers is the first stage. The second stage involves identifying and isolating the number's PVPs and showing the total of their sum. This stage is frequently described as "expanded form" in first and second grades in US schools, but rarely utilised (Howe, 2019). As such, many American third and fourth grade students were not able to think of the decimal number system in terms of the second stage structure (Kamii, 1986). Using the common core standards in US education as an example, in third grade it is possible to advance to stage three, but only after multiplication is introduced. This stage is crucial for a conceptual understanding of base-10 addition and subtraction. However, Tannheiser (2009) reports that in the US, the large majority of students arrive in college without understanding the third stage.

The following two stages refine stage three by making the multiplicative structure of the base-10 components more explicit. Stage four shows the base-10 units as products of multiple factors/powers of ten, necessitating a high level of multiplication proficiency. The last step employs exponential notation to represent the powers in a more compact manner, and hence would not be available until this notation is studied in 6th grade according to the common core standards in US education. At the end of the operation, 4×10^2 represents the hundreds place (4 hundreds or 400), 7×10^1 represents the tens place (7 tens or 70), and 2×10^0 represents the units place (2 units or 2).

As mentioned above, Tannheiser (2009) reports that of the US students arriving in college that participated in her study, 80% had only a limited understanding of the place value system. One reason for this can be found in the idiosyncrasies of the English counting sequence. The fewer idiosyncrasies that occur within a counting sequence, the easier it is for children to grasp its organisation. This not only speeds up the learning process of the counting sequence, but it can also enhance the development of arithmetic concepts by making it easier to see multiplicative and additive relations between numbers. A further advantage of early mathematical learning is the fact that a more transparent counting sequence is also easier to match with the common notation of Arabic numerals. The lack of a clear pattern in English number names can make grasping a number's place value very difficult for students (Morin & Franks, 2009). Number words form a special category within languages and their internal organisation as well as its influence on learning will be further explored in the next section.

3.3. Numerical Vocabulary

Numerical vocabulary is distinct from the rest of the language in several ways. Specifically counting or number words diverge from other lexical elements in four ways. As Wiese (2003, p. 70) explains,

> [number words] form a **progression**, this progression is **infinite**, its elements are **well distinguished by their respective phonological representation**, and they are **non-referential**. The first three of these features establish the numerical character of counting words; the last one, their non-referentiality, confines them to the function of numbers. (my emphasis)

The first two elements are proposed as distinctive properties of numbers, being ordered as a progression and being infinite, differentiate counting sequences from other lexical domains, such as words for plants. The words for 'tree', 'flower', 'grass', and 'cactus' are not a sequence like ONE, TWO, THREE, and FOUR is. Furthermore, due to recursiveness, it is certain that TWENTY-THOUSAND-FIVE-HUNDRED-SIXTY-ONE has a successor in TWENTY-THOUSAND-FIVE-HUNDRED-SIXTY-TWO, which in turn has a successor *ad infinitum*, but there is no method or system that gives the next plant name for any plant name that one can conceive of.

Recursive rules that regulate the creation of counting words provide the potential for infiniteness in any counting sequence. These rules take a set of counting words as input and produce a set of new counting words as output. They usually have additive character, such as the rule that creates SEVENTY-THREE from SEVENTY and THREE, or multiplicative character, such as the rule that creates FIVE MILLION by combining a counting word like FIVE with a base word like MILLION. As a result, the fact that counting sequences are endless does not entail that they contain an unlimited number of primitive elements. This means

that there are morphosyntactic rules that produce an infinite number of counting words based on a finite set of lexical primitives and a stock of bases, such as -TY, HUNDRED, THOUSAND, MILLION, and other components that are utilised in combination with both primitive and complex counting words. Every item in this description that acts as one part of the counting sequence is referred to as a counting word. As a result, there exist morphologically simple counting words like SEVEN and more complicated ones like TWENTY-THOUSAND-FIVE-HUNDRED-SIXTY-TWO (see Hurford, 2011, for an in-depth account of phrase structure rules underlying the generation of complex counting words within generative grammar).

As a result, counting words from an infinite set, and the elements within that set, have an inner structure that serves as the foundation for their sequential order. To illustrate, the following are the rules for English:

- The primitive elements ONE to NINE are the basic elements needed in order to generate the English counting sequence.
- The order of these items can be specified directly via enumeration: ONE, TWO, THREE, ... NINE because this set is finite and small.
- This will form the basis for defining the rules and will be used to produce complex counting words of various types.
- These can be defined as distinct classes within the counting sequence, i.e. the elements of the Teen-class (counting words from TEN to NINETEEN), the Ty-class (TWENTY to NINETY-NINE), the Hundred-class (ONE HUNDRED to NINE HUNDRED AND NINETY-NINE), the Thousand-class (ONE THOUSAND to NINETY-NINE THOUSAND NINE HUNDRED AND NINTEY-NINE), the MILLION-Class (from ONE MILLION upwards), and so on.

The classes may now be sorted in order of decreasing complexity, with simple counting words being the lowest, followed by the base classes, such as Teen-class, Ty-class, Hundred-class, and so on. Multiplicative rules combine a base (-TY, HUNDRED, THOUSAND…) with a counting word that acts as a multiplier for each class. The rule creates SEVENTY from the base -TY and the multiplier SEVEN in the Ty-class, SEVEN HUNDRED from the base HUNDRED and multiplier SEVEN in the Hundred-class, and so on. The exception is the Teen-class, which has a base, TEEN, but which does not function as a multiplier. This is because there is no “oneteen” but instead “ten”, and the ELEMENT that should be “twoteen” is TWENTY, which is already a Ty-class element.

There are also additive rules in English which take counting words like SEVEN HUNDRED as input and combine them with parts from lower classes to

produce counting words like SEVEN-HUNDRED-AND-THIRTY. Again, the Teen-class deviates slightly in that its initial three components (TEN, ELEVEN, and TWELVE) are linguistically opaque, in the sense that they do not reveal the constituent structure that links them to other counting words. This is in opposition to, for example, SIXTEEN, which is formed from SIX and the base TEEN.

As a result, these elements could be declared as primitives by enumeration in the same way that the elements' original class (ONE to NINE) is defined. Elements of the same class are ordered by the ranks of their multipliers. For example, FORTY comes before EIGHTY because FOUR comes before EIGHT. Additionally, elements with the same multiplier are ordered by their increments, for example, FORTY-FIVE comes before FORTY-NINE because FIVE comes before NINE. The multiplier must always be an element of a lower class when modifying the multiplicative elements of the Ty-, Hundred-, and Thousand-classes. For example, TWO THOUSAND and TWO HUNDRED THOUSAND are permissible, but *TWO THOUSAND HUNDRED or *TWO MILLION THOUSAND are not.

This constraint does not apply to the highest classes, which are open and allow the sequence to be endless. These open classes contain combinations like FIVE MILLION MILLION (= 5×10^{12} or 5,000,000,000,000) or THREE BILLION BILLION BILLION (= 3×10^{27}) starting with the Million-class upwards. This option includes recursion, because it holds for multiplicative elements of the Million and above classes that, combining a counting word with a [high-class-word], will result in a new counting word. This term then becomes the foundation

for a whole new set of additive elements. There will never be an end to such multiplicative elements and no limit to new input words since the output of this rule is again a counting word and therefore qualifies as a new input word. Counting sequences are limitless due to the recursivity of their generating principles.

3.3.1. Inconsistencies

The rules laid out above provide an idealised version of the English counting sequences because the additive and multiplicative character of the rules controlling the formation of counting sequences is not always transparent to the same degree. A counting sequence can also have components that are more or less opaque as a result of language change. For example, instead of TWENTY, THIRTY, FORTY, and FIFTY, the idealised rules would yield elements like TWO-TY, THREE-TY, FOUR-TY, and FIVE-TY. Additionally, because counting words denoting lower-end numbers are used more frequently than those denoting higher ones, they are more prone to idiosyncrasies.

In general, less frequently used words have the tendency to become more regular, because the irregular form they might have had originally is not encountered frequently enough to be memorised by future generations learning the language. Therefore, irregular terms that are rarely used tend to become increasingly regularised over time (Lieberman et al., 2007). As an example, the components ELEVEN and TWELVE are now opaque, meaning that there is no underlying structure available that shows their sequential location in the same manner that EIGHTEEN fits between SEVENTEEN and NINETEEN.

ELEVEN and TWELVE are, however, derived from the Old English words *endleofan* and *twelf*. These lexical items can be traced back to Proto-Germanic words with internal structure, as the Gothic *ain* + *lif* and *twa* + *lif* would indicate. *Lif* comes from a word meaning 'to leave', and so *ainlif* means 'one left (after ten)' and *twalif* means 'two left (after ten)'. This implies an now obsolete upper limit of ten in the counting sequence (Davis, 1998; von Mengden, 2005). The same derivation holds true for the German and Danish counting words for eleven (German *elf*, Danish, *elleve*) and twelve (German *zwölf*, Danish *tolv*).

In addition to idiosyncratic words, it is also possible for a language to have counting words that are fully transparent in terms of their internal structure, but which are created by an idiosyncratic rule. Using a subtractive rule alongside an additive rule is an example of this. We see this in Latin for example:

- In Latin, the two counting words that come before VINGINTI (twenty) are DUO-DE-VINGINTI ('two from twenty' = eighteen) and UN-DE-VINGINTI ('one from twenty' = nineteen). This is a subtractive rule.

- However, according to the normal additive method that underpins the Latin counting sequence, these words should in fact be defined as the elements that appear in the eighth and ninth positions after ten, akin to English EIGHTEEN and NINETEEN.

Over-counting, a concept coined by Menninger (1958) and later used by Hurford (1975) and others in studies on numerical systems, is comparable to this. In an over-counting system, a value is expressed in relation to a higher parameter known as the *augend*. Over-counting literally expresses the number of units

counted towards the augend: it is a counting-word based on over-counting which gives its position within the next block as defined by the respective multiplier. In this system, SIXTEEN would be produced as 'the sixth element within the block ending at twent'.

Over-counting systems are extremely rare typologically, having only been documented in a few languages spoken in a few places around the world such as in Mayan languages spoken in Mexico (Aulie 1957), in the variety of Ainu spoken on Sakhalin Island in north-east Asia, and in the pre-Christian Germanic north of Europe (Menninger, 1979, 2013). Mazaudon (2009) has reported a limited overcounting pattern in the Dzongkha language of Bhutan, as well as traces of such a system in other Tibetic languages. It appears that this latter example almost certainly developed independently of the over-counting numeral systems in the languages of Nagaland in northeastern India, based on current knowledge of likely migration routes of speakers across the Tibetan plateau, among other factors (Coupe, 2012). A good example of over-counting may be

viewed in Finnish (Menninger, 1979, 2013).

A counting sequence can also include sub-sequences that use different multipliers from the rest of the system. French provides a good example of this phenomenon:

- In Finnish, ‘sixteen’ becomes KUUSI-TOISTA, which means 'six-second’, that is, the *sixth* element inside the *second* decade.
- This over-counting rule is used in Finnish exclusively for the counting words sequence between ‘eleven’ and ‘nineteen’.
- It is not used for counting words up to ten nor for words in higher classes from twenty.
- Many deviations from a counting sequence's general building structure, such as this one, emerge from and are limited to very small subsystems of counting words produced by non-standard rules.

- In French, both QUATRE-VINGTS (“four-twenties” = 80) and QUATRE-VINGT-DIX (“four-twenty-ten” = 90) use multiples of twenty rather than the decimal system, which provides the basis for the majority of the French counting sequence.

- Consequently, the numbers 70–79 and 90–99 also use the teen numbers in place of the numbers of the main cycle in the unit position SOIXANTE-DIX (“sixty-ten” =70), SOIXANTE-ET-ONZE (“sixty-and-eleven” = 71); 76: SOIXANTE-DIX-SEPT (“sixty-ten-seven” = 76), etc, and QUATRE-VINGT-ONZE (“four-twenty-eleven” =91), QUATRE-VINGT-TREIZE (“four-twenty-thirteen”= 93), QUATRE-VINGT-DIX-NEUF (“four-twenty-ten-nine” = 99).

Having looked at the more general nature of inconsistencies in number naming, we now turn to inversion, a specific inconsistency that will play an important role later in the discussion of the articles identified in the scoping review.

3.3.2. Inversion

A deviating subsystem may also be the result of rules that generate a non-standard order for the parts of certain complex counting terms. In English, the names for THIRTEEN to NINETEEN form a subsystem, in which, instead of utilizing the standard rules, parts of additive counting terms are inverted. The standard rules would instead generate TEN-THREE etc., which would correspond to TWENTY-THREE, THIRTY-THREE, and so on. In German and Arabic counting sequences, this type of subsystem is more extensive and covers

the whole sequence from “thirteen” to “ninety-nine”. In these languages, additive counting words result in:

- ‘one-and-twenty' (German: ZWEIUNDZWANZIG; Arabic: 'واثني ن و عش_رون')
- ‘three-and-twenty’ (German: DREIUNDZWANZIG; Arabic: 'وعش_رون ثلا_ثة')
- ‘nine-and-ninety’ (German: *NEUNUNDNEUNZIG*; Arabic: 'تسع وتسعون').

Compare with Italian, which uses this subsystem solely for a section of the Teen-class:

- Employed in the counting terms 11-16, UN-DICI (eleven), DO-DICI (twelve), TRE-DICI (thirteen), QUATTOR-DICI (fourteen), QUIN-DICI (fifteen), and SE-DICI (sixteen).
- Counting words from seventeen upwards are in standard order, but sometimes do use the and-infix (DICI-as-SETTE ‘ten-and-seven’) and sometimes not (DICI-OTTO "ten-eight").

In Spanish, the numbers 11-15 all have unique names that must be learned by heart, like the numbers 1-10. Nonetheless, they follow an inverted pattern, with these numbers’ Teen-marker in second place, however the first syllable of some only vaguely corresponds with the numbers 1 – 5.

- 11 = ON-CE (“one-[Teen-marker]), while 1 = UNO; 12 = DO-CE (“two-[Teen-marker]”), 2 = DOS; 13 = TRE-CE (“three-[Teen-marker]”), 3 = TRES; 14 = CATOR-CE (“four-[Teen-marker]”), 4 = CUATRO); 15 = QUIN-CE (“five-[Teen-marker]”), 5 = CINCO.

- The numbers 16 – 19 then become regular in their order and the indicator of the Teen-class corresponds to the actual word for 10 (DIEZ).

- From then on, all numbers from 16 to 99 use the and-infix, apart from numbers 16 – 29, which contract to form what would be the regular form DIEZ-y-SEIS to DIECISÉIS (“ten-and-six”), DIEZ-y-SIETE to DIECISIETE (ten-and-seven), DIEZ-y-OCHO to DIECIOCHO (ten-and-eight), and so on.

- Interestingly, the name for the numeral 20 in Spanish is VIENTE, which also does not correspond to the name for 2, DOS, and does not use the regular Ten-class marker “-TA”, and has therefore also to be learned as a distinct element.

- Numbers from 30 – 99 follow the regular form, e.g. 31 = TRIENTA Y UNO, 55 = CINCUENTA Y CINCO, etc.

These examples show that departures from a counting sequence's general structure are a common in languages. What matters is that this is limited to a few locations and a small number of elements, so these sub-sequences or individual elements may be learned as distinctive components.

3.3.3. Homophony and Non-referentiality

Homophony occurs frequently in all languages, but it is completely absent within any given system of counting words. Each counting word appears only once in its counting sequence, so no two counting words have the same phonological expression. There will never be a term like “FIFREE” that may mean either “fifteen” or “three” owing for instance to language change and leading to queries like "Do you mean “FIFREE” as in “TWO-FIFREE-FOUR” or “FOURTEEN-FIFREE-SIXTEEN”? This is thought to be due to each counting word functioning as a number in and of itself (Goddard, 1961). On the other hand, this does not mean that multiple words may not be used to represent the same place in the counting sequence. In English, “ONE THOUSAND FIVE HUNDRED” and “FIFTEEN-HUNDRED” are two salient examples.

The last unique feature of number words is that they are a distinct subcategory of language in and of themselves. Unlike other words, counting words have no meaning and do not relate to anything in the outside world. This is because counting words are tools employed in number assignment and do not require any referentiality. Their non-referentiality is the result of all of the above characteristics since numbers should be distinct, infinite, and sequential: “they should have ‘number’ attributes – but nothing else” (Wiese, 2003, p. 80). Furthermore, the fact that a counting sequence may be imported into a language in its entirety and may coexist alongside native sequences (such as the ancient Chinese system still used in Korean and Japanese) emphasises the unique standing of number words within a linguistic system, where number words are intricately linked to one another, forming a complex interconnected

progression. They are also distinguished from other linguistic domains by their non-referential nature, which is another one of the four characteristics identified by Wiese (2003) as distinguishing counting words from other lexical items. Next, I will discuss the number naming system of two languages, Danish and Irish. Both languages are highly opaque and feature several inconsistencies that serve to illustrate many of the kinds of irregularities that exist in one way or another in other languages, including those examined in the scoping and systematic reviews conducted later in this thesis.

3.4. The Example of Danish

Among European languages, the Danish number naming system stands out for containing various subsystems and deviations from the basic structure of the counting sequence. Danish number names are highly complex, stemming from historical base 12 and base 20 number systems, as we have seen briefly above. Its number names are described in the *Great Danish Encyclopaedia* as "very old and reflecting a number concept that is primitive in relation to numerical thinking" (Talord, 2009-10, translation via Ejersbo & Misfeldt, 2011, p. 1). Whereas the three Scandinavian languages of Swedish, Norwegian, and Danish are mutually intelligible, the number naming systems in Sweden and Norway are comparable to those in English, while the Danish system is fundamentally different (Ejersbo & Misfeldt, 2011).

Ejersbo and Misfeld (2011) compared the first 100 numbers in Danish number naming system with the English and Japanese systems, focusing on two main issues: (1) number words that do not follow an internal logic and therefore must be memorised and (2) the overall regularity of the spoken system, i.e. how

closely the spoken system corresponds to the written base-10 system of Arabic numerals. The results of the comparison are presented below in Table 3.1 for reference.

Table 3.1: Number names between 1 to 100 in Danish, English, and Japanese, combined with the calculations underlying their names (from Ejersbo & Misfeldt, 2011)

		Explanation	English	Explanation	Japanese (romaji)	Explanation
1	*En*	1	one	1	*Ichi*	1
2	*To*	2	two	2	*Ni*	2
3	*Tre*	3	three	3	*san*	3
4	*Fire*	4	four	4	*shi*	4
5	*Fem*	5	five	5	*go*	5
6	*Seks*	6	six	6	*roku*	6
7		7	seven	7	*sichi*	7
8	*Otte*	8	eight	8	*hachi*	8
9	*Ni*	9	nine	9	*kyu*	9
10	*Ti*	10	ten	10	*juu*	10
11	*Elleve*	11	eleven	11	*juu-ichi*	10 + 1
12	*Tolv*	12	twelve	12	*juu-ni*	10 + 2
13	*Tretten*	13	thirteen	13	*juu-san*	10 + 3
14	*Fjorten*	14	fourteen	14	*juu-shi*	10 + 4
15	*Femten*	15	fifteen	15	*juu-go*	10 + 5
16	*Seksten*	16	sixteen	16	*juu-roku*	10 + 6
17	*Sytten*	17	seventeen	17	*juu-sichi*	10 + 7
18	*Atten*	18	eighteen	18	*juu-hachi*	10 + 8
19	*Nitten*	19	nineteen	19	*juu-kyu*	10 + 9
20	*Tyve*	20	twenty	20	*ni-juu*	2x10
21	*Enogtyve*	1 + 20	twenty-one	20 + 1	*ni-juu-ichi*	2x10 + 1
30	*Tredive*	30	thirty	3x10 or 30	*san-juu*	3x10
32	*Toogtredive*	2 + 30	thirty-two	30 + 2	*san-juu-ni*	3x10 + 2
40	*Fyrre*	40	forty	4x10 or 40	*si-juu*	4x10
43	*Treogfyrre*	3 + 40	forty-three	4x10 + 3	*si-juu-san*	4x10 + 3
50	*Halvtreds*	50 (2.5x20)	fifty	5x10	*go-juu*	5x10
54	*fireoghalvtreds*	4 + 50	fifty-four	5x10 + 4	*go-juu-shi*	5x10 + 4
60	*Tres*	60 (3x20)	sixty	6x10	*roku-juu*	6x10
65	*Femogtres*	5 + 60	sixty-five	6x10 + 5	*roku-juu-go*	6x10 + 5
70	*Halvfjers*	70 (3.5x20)	seventy	7x10	*sichi-juu*	7x10
76	*seksoghalvfjers*	6 +70	seventy-six	7x10 + 6	*sichi-juu-roku*	7x10 +6
80	*Firs*	80 (4x20)	eighty	8x10	*hachi-juu*	8x10
87	*syvog firs*	7 + 80	eighty-seven	8x10 + 7	*hachi-juu-sichi*	8x10+7
90	*Halvfems*	90 (4.5x20)	ninety	9x10	*kyu-juu*	9x10
98	*otteoghalvfems*	8 + 90	ninety-eight	9x10 + 8	*kyu-juu-hachi*	9x10 + 8
100	*Hundrede*	100	hundred	100	*hyaku*	100

Ejersbo and Misfeldt (2011) noted that the first 100 number words in Danish include 28 basic distinct names for the numbers 1-9, 10-20, 30, 40, 50, 60, 70, 80, and 90, whereas written base-10 only has 10 separate signs integrated in a system. To count to 100, Danish children must remember 28 different words,

whereas Japanese children only need to learn 11 different terms by heart. In English, the number depends on whether you count the different iterations of ten (-ten, -teen, and -ty) as separate entities or variants of the same word. More precisely, there are 13 words to memorise if all three are accepted as essentially the same word, 20 to learn if -teen is accepted as the same as ten, and 28 if all three are deemed different entities.

As in German and Arabic, the inversion-subsystem also encompasses Danish number names from 13 to 99, with 'and' added between 21 and 99, e.g., 13 = TRE-TEN (three-ten), 23 = TRE-**og**-TYVE (three-**and**-twenty), and so on. As in German and English, Danish also uses opaque words for 11 (*elleve*) and 12 (*tolv*) that do not fit into the - ten ending structure followed by the names for numbers 13-19. This is a remnant of an old system with an upper limit of 10 (Ejersbo & Misfeldt, 2011). All numbers from 13-19 in Danish end in *-ten* (-teen). However the first syllables of these names do not sound like their corresponding numbers 3-9 and neither are the Danish -ty class number names from 20-90 connected to their counterparts 2-9 either. For example 8 = *otte*, 18 = *atten*, 80 = *firs*. In comparison to English, this makes Danish number names substantially more difficult for the mind to systematise and therefore harder to learn (Rischel, 2012).

The fact that, starting from 40, Danish number names follow a base twenty number naming system, rather than the regular base-10, is another very confounding factor for Danish children when learning to count. While, as we have seen, French also possesses a similar subsystem, the Danish one is more complex, and the number naming itself is more opaque. The names of various -ty class numbers in Danish have been simplified in recent years, but their

etymology still very markedly a base twenty system. Additionally, from 40 to 90, the Danish -ty class number names also use fractions as multipliers. The fractional system implemented in the names of these numbers takes the first half is ½ (0.5), second half is 1½ (1.5), third half is 2½ (2.5), and so on. The base twenty system along with the use of fractions as multipliers means the numbers 40-90 are presented as follows (Rischel, 2012; Konig & Van der Auwera, 2013):

- Fyrre (40), short for fyrretyve, is an exception that derives from the Old Norse word fyritiugh, which means 'four tens'. The literal translation of the word fyrretyve is 'four twenty'.

- Halvtreds (50) short for halvtredje-sinds-tyve, literally meaning 'third half times twenty' or 'two twenties plus half of the third twenty' (i.e. 2½ x 20).
- Tres (60), short for tre-sinds-tyve, means 'three times twenty' (i.e. 3 x 20).
- Halvfjerds (70), short for halvfjerds-sinds-tyve, means 'fourth half times twenty' or 'three twenty plus half of the fourth twenty' (i.e. 3½ x 20).
- Firs (80), short for firsindstyve or fire-sind-styve, means 'four times twenty' (i.e. 4 x 20).
- Halvfems (90), short for halvfemsindstyve or halv-fem-sind-styve, means 'fifth half times twenty' or 'four twenty plus half of the fifth twenty' (i.e. 4½ x 20).

The place value system of Arabic numerals is greatly incongruent with Danish number names, which are very complex. When asked to say the name of a written numeral, such as 63, Danish first graders frequently recite a ten, twenty, thirty rhyme while counting up to the sixth finger, then they stop and know the

word. This could be interpreted as the two simultaneous processes being in conflict, with semantic processing requiring too much attention and taking longer, making it impossible to multitask and finish both processes at the same time (Baddeley, 2009). To put it another way, the logic of the base-10 system gets lost within the Danish number naming convention, and number names must be memorised. As a result, learning to count, combining written and spoken number systems, takes substantially longer for Danish children (Ejersbo & Misfeldt, 2011).

A less severe lack of a clear pattern can also be found in English number names, and this might make it difficult for pupils to comprehend place value (Morin & Franks, 2009). An absolutely regular English counting sequence, as in Japanese, such as ten, ten-one, ten-two, ten-three etc., two-ten, two-ten-one, two-ten-two, etc., combined with the removal of all opaque items such as eleven, twelve, thirteen, and twenty, would make it much easier to understand for instance that:

- 10 + 2 = 12 ('ten-two')
- 13 ('ten-three') - 3 = 10
- 10 x 2 = 20 ('two-ten').

3.5. The Example of Irish

An entirely different kind of number system can be found in the Irish language. The Irish language (*Gaeilge*), a member of the Goidelic branch of the Celtic language family, demonstrates a very rich and layered number and counting system that combines elements of grammatical inflection, collective categorization, and remnants of archaic numerical bases. Numerals in Irish are broadly categorized into basic cardinals, ordinal forms, and collective numerals specifically for humans, each of which triggers distinct syntactic and

morphological behaviours. Cardinal numbers from one to ten (Table 3.2) follow a relatively straightforward structure, though they diverge from English in form and grammatical behaviour, such as being used after the particle "*a*".

Table 3.2: Irish Cardinal Numbers (1–10)

Number	Irish	Number	Irish
1	*a haon*	6	*a sé*
2	*a dó*	7	*a seacht*
3	*a trí*	8	*a hocht*
4	*a ceathair*	9	*a naoi*
5	*a cúig*	10	*a deich*

From eleven onwards, traditional Irish used inversion, e.g., 13 = *a trí déag* (three-ten), however this changed due to the standardisation of Irish education, especially post-1950s. In modern Irish non-inverted number names tend to be used. Similarly, the vigesimal or base-20 system was widely used in Irish before the mid 1900s, until base-10 was formally promoted in modern standardised usage. The only number name that still derives from the old base is 40, namely *daichead* = 2×20 (Ó Siadhail, 1989; Mac Mathúna, 2006).

One of the most typologically unique features of Irish is its use of collective numerals exclusively for counting humans. These forms (e.g., *triúr* for three people, *cúigear* for five) are listed in Table 3.3 and typically appear with singular genitive nouns (e.g., *triúr ban* – "three women") regardless of the actual quantity being referenced. This system is still robustly used in oral speech, particularly in Gaeltacht communities (Ó Baoill, 2009).

Table 3.3: Collective Numerals for Humans

Number	Human Form	Example Phrase
2	*Beirt*	*beirt bhan* (two women)
3	*Triúr*	*triúr buachaillí* (three boys)
4	*Ceathrar*	*ceathrar fear* (four men)
5	*Cúigear*	*cúigear páistí* (five children)

Furthermore, Irish numerals interact intricately with the grammatical system, regularly causing either lenition or eclipsis on the following noun depending on the numeral and noun class (Table 3.4). For instance, *sé chat* ("six cats") involves lenition, while *seacht gcapall* ("seven horses") involves eclipsis, demonstrating the interplay between phonology and morphosyntax (Ní Chiosáin & Tallerman, 2011). When speaking about more than 10 items, the item in question is placed between the first and second numbers e.g. I'm 37 years old = *Tá me 7 mbliana is 30 d'aois* (I am 7 years and 30 old).

Table 3.4: Mutation Effects Caused by Numerals

Mutation Type	Example
Lenition	*sé chat* – "six cats" (cat → *chat*)
Eclipsis	*seacht gcapall* – "seven horses" (*capall* → *gcapall*)

Ordinal numbers in Irish follow a consistent prefix-based system, typically with the definite article "*an*" followed by the ordinal (e.g., *an tríú lá* – "the third day") and usually trigger lenition on the subsequent noun (Table 3.5).

Table 3.5: Ordinal Numbers in Irish

Number	Ordinal Form	Example
1st	*an chéad*	*an chéad lá* – "the first day"
2nd	*an dara*	*an dara teach* – "the second house"
3rd	*an tríú*	*an tríú bliain* – "the third year"
4th	*an ceathrú*	*an ceathrú uair* – "the fourth time"

While this system remains largely stable, modern usage is marked by a tendency toward simplification and standardization, particularly in educational contexts and official communication (Ó Murchú, 1985; Ó Siadhail, 1989; Mac Mathúna, 2006; Hickey, 2011). Nonetheless, traditional vigesimal and collective structures continue to persist in dialects and informal speech, creating a dual-register system where decimal and base-20 numerals coexist (Ó Siadhail, 1989; Mac Mathúna, 2006).

3.6. Conclusion

Having looked at how both Danish and Irish as a means of illustrating how number naming systems can differ, it is clear how linguistic differences in number naming could affect young children's formal numerosity acquisition in important ways. The assertion of this chapter is that number words are indeed special. The sections above have examined the emergence of the language-related components of numerical capabilities, with an emphasis on the significance of studying linguistic variations in number naming and within number naming systems, such as inversion. The decimal system and other key aspects of the special nature of number words were explored, such as the fact that number words form an infinite progression with elements that are non-referential and well distinguished by their respective phonological representation. In the following chapter the process of acquiring number words and concepts in children will be introduced.

Chapter 4: Acquisition of Number Words and Concepts in Children

4.1. Introduction

This chapter demonstrates how a number naming system's transparency promotes early learning in young children and how discrepancies can hamper counting-word acquisition. The cardinality principle and the idea of number-knowers, which is the incremental understanding that the last number in a counting sequence indicates the entire number of objects in a set, will be introduced. The importance of semantic singular/plural markers inside a language on young children's comprehension of cardinality is examined. Dimensionality in connection with numeration systems is introduced, and by using the often-used metaphor of the mental number line, the mental representation of numbers is demonstrated. This is followed by an examination of how numerical estimation skills develop, how representations on the mental number line manifest, and finally the impact of language on mental number line representation.

4.2. Transparency and Learning

To understand the challenges associated with learning number words in one's first language, it is necessary to understand the significance of linguistic transparency in number naming which can affect mathematical comprehension, even beyond the basic stages of counting and number comprehension. Returning to the base-10 system described in the previous chapter, according to Lam and Ang (2004), it is interesting to note that the decimal system as we know it today is a textual version of the ancient Chinese counting rods technique of

computation (Sun, 2015). Research shows that children learning to count using the Chinese sequence have an easier time understanding the fundamentals of base-10 place-value notation than their English-speaking American peers. When asked to make the number 25 out of unit cubes and bars of ten, Chinese pre-school children quickly chose two bars of ten and five units, which indicates that they are familiar with base-10. At the same age, American children responded differently. The majority laboriously counted out twenty-five units, without making use of the ten-unit shortcut. Furthermore, if the English-speaking children were given a twenty-unit bar, they would utilise it more frequently than two ten-unit bars. This indicates that they appeared to be preoccupied with the surface form of the term 'twenty-five', whereas the Chinese children seemed to have understood the underlying base-ten structure. (Miller and Stigler, 1987).

In Japan, and Korea, counting sequences that are based on the ancient Chinese one that continues to be used, in addition to indigenous counting sequences. It has been reported in a number of studies (e.g., Fuson & Kwon, 1992; Miller et al., 1995; Miura et al., 1994; Saxton & Towse, 1998) with Chinese, Korean, and Japanese compared with US-American, French, and Swedish first graders, that the Asian first-graders who used the Chinese system had a significantly better understanding of the base-ten structure of their number sequence and of the place value system underlying the use of digits in Arabic numerals. According to Miura et al. (1994), first-graders in the US, France, and Sweden have to learn how to deal with irregular number names that don't give any phonetic indications to their base-ten composition, like the English *eleven* and *twelve* or the French *onze* and *douze*. Children in these cultures frequently first view 11 or 12 as distinct,

unitary values rather than a combination of "*ten + one*" and "*ten + two*" due to the arbitrary nature of these Western number labels. As a result, Asian students are more likely to employ the canonical base-ten strategy (grouping by tens) when asked to represent numbers using base-ten blocks, while their Western counterparts are much more likely to count out individual units, indicating a less developed understanding of the decimal structure.

Researchers Miura and Okamoto (Miura et al., 1988, 1994; Miura & Okamoto, 1989) concluded that the structures of different counting sequences may have a major influence on how children represent numbers in their minds. In other words, whereas base-ten is a transparent notion in those languages that have adopted the Chinese counting sequence, it is considerably more difficult for children using counting systems from European languages to comprehend. The examples of irregularities drawn from Danish and Irish in the previous chapter make it clear why this would be the case. While Towse and Saxton (1998) suggest that Muira and Okamoto's findings are influenced in part by their experimental design, other research has found comparable outcomes for Chinese and Korean pre-schoolers compared to British and American pre-schoolers (Dowker & Li, 2019; Fuson & Li, 2009; Ho & Fuson, 1998; Miller & Zhu, 1991; Song & Ginsburg, 1988). According to longitudinal data from English- speaking and Spanish-speaking children, the introduction of a programme incorporating explicit number naming at the pre-school level enhances children's number sense and place-value recognition in primary school (Magargee & Beauford, 2016). This topic will be addressed in more detail later in this thesis.

4.3. The Acquisition of Counting Words by Children

Given the non-referentiality of number words, a question arises about how children understand that a number word refers to a number rather than to an object or an attribute of an object. Much of what infants hear in regard to numbers is barely usable in terms of learning (Gelman & Gallistel, 1986). In many countries, children are most likely to encounter number words as part of the routine of verbal counting, although these counting exercises are usually done without any objects present, as in rhymes and routines like "one, two, three, go" (Durkin et al. 1986). The child has no indication that these words correspond to numbers rather than being strings of nonsensical words like "eeny, meeny, miny, moe" or the letters of the alphabet A, B, C.

Of course, children do occasionally witness the counting of entities, whereby each number word is assigned to an individual object in a one-to-one relation, such that one item is labelled with 'one', another with 'two' and so on (Gelman & Gallistel, 1986). Nonetheless, as the transcript below from a mother-child interaction by Durkin et al. (1986, 281) illustrates, even then it may be rather perplexing for a child. The mother changes from counting fingers, i.e. assigning them counting words non-referentially, to identifying fingers, i.e. assigning them nouns referentially. This discussion took place after the mother and daughter (24 months) had counted the four cameras in the studio.

Table 4.1 Interaction between mother and child (Durkin et al., 1986, p. 281)

Mother (to Lucy):	One	*(raises thumb)*
Lucy:	One	*(waves arm)*
Mother:	Two	*(raises thumb and index finger)*
Lucy:	Two	*(waves arm)*
Mother:	Three	*(raises thumb, index and middle finger)*
Lucy:	Sree	*(waves arm)*
Mother:	Four	*(raises thumb, index, middle and ring finger)*
Lucy:	Four	*(hands in lap, smiling)*
Mother:	What's that?	*(raises index finger)*
Lucy:	Two	
Mother:	Finger	
Lucy:	Finger	
Mother:	And what's that?	*(raises thumb)*
Lucy:	One two three ror	*('one' pronounced emphatically)*
Mother:	Thumb	

Entities that are counted are often objects and, taking into account children's propensity to treat words as object names (Huttenlocher & Smiley, 1987), it might be especially difficult for them to learn that the number words do not describe the individual items being counted. Another possibility might be that children would think that such words were demonstrative pronouns, akin to ‘this’ or ‘that’. This would be a reasonable mistake to make, but it is one that children somehow never make (Bloom, 2002). Children learn to count early, and 2-year-olds can usually count up to three or more (Gelman & Gallistel, 1986). Once they have learned to count objects, they can also count actions and sounds (Wynn, 1990). Around the age of four, children start to understand some of the generative rules that underpin the language of numbers (Fuson et al. 1982; Siegler & Robinson, 1982).

4.4. The Cardinality Principle

Studies have found that children comprehend the meaning and idea of numbers progressively. According to researchers such as Schaeffer, Eggleston and Scott (1974) and Gelman and Gallistel (1986), correct counting follows a set of rules. Such rules include reciting the counting sequence in the same order every time, pointing to one and only one object with every number word said, pointing to each and every object without skipping or double counting any of them, and, lastly, understanding that the last number stated denotes the complete number of objects in the whole set. That last rule was called the cardinal principle by Gelman and Gallistel (1986) and the cardinality rule by Schaeffer et al. (1974), and it is the only one, that involves something that a child knows rather than something they do, as discussed earlier in this thesis.

Measuring what children know, however, is more difficult than observing what they do. While the difference between knowledge and action is true for all behaviour to some extent, a gap in children's counting behaviour has been used as evidence for and against the notion that children possess concepts of natural numbers (Wiese, 2003). It is worth noting that these natural number concepts are mental representations of exact numerical quantities, such as five or eight, and are not used to indicate knowledge of mathematical entities termed ‘natural numbers’.

In practice, this means children learn to count long before they understand that the last number they reach reveals the cardinal number of the entire set. To put it another way, infants first acquire procedural knowledge, i.e., the rules described above, and then gain conceptual knowledge, i.e., the cardinal-rule, with the

aforementioned gap between action and knowledge (Freeman, et al., 2000; Frye, et al., 1989; Fuson, 1988; Miller, et al. 1995; Slaughter, et al., 2011). At around 3-and-a-half years old, English-speaking (Wynn, 1990) and Chinese-speaking (Jing, 1987) children start using the last word from the count to indicate the cardinality of a whole set.

The counting rules first appear to constitute a placeholder structure, implying that number and counting concepts are acquired through a type of learning known as conceptual-role bootstrapping (Carey, 2009). The learner first acquires a placeholder structure consisting of a collection of symbols in fixed relation to each other as part of the conceptual-role bootstrapping process. Individual symbols begin as placeholders, characterised by their location within the framework. They will only later be imbued with meaning and only become fully formed concepts over time.

The so-called 'Give-N' or 'Give-A-Number' task is one of the commonly used experiments that illustrates this gap. Le Corre et al. (2006) utilised the following design: a child was given a set of objects (for example, a bowl with 15 miniature plastic apples) and instructed to deliver a specific number of the objects to a puppet (e.g. 'give six apples to Kermit'). Many children between the ages of two and four are unable to provide the proper quantity of apples in the Give-N tasks, despite the fact that they can count a row of ten apples correctly without mistakes. Instead of counting to produce the required number, children typically just take a handful of items. Even when specifically prompted with questions like "Can you count to make sure you gave Kermit six apples?" or "But Kermit wants six apples – can you make it so that there are six apples?", they fail to apply the result of counting to accomplish the task. The Give-N task has been used in many studies

(e.g. von Hanxleden & Kennedy, 1994; Negen et al., 2012; Marchand & Barner, 2020; Wege et al., 2022; Connor et al., 2024; Le et al., 2024) and results show that children's learning of number concepts progresses through distinctive and predictable stages in terms of performance.

4.5. The Concept of Number-Knowers

These number learning stages are often referred to as number-knower levels (Condry & Spelke, 2008; Le Corre et al., 2006; Le Corre & Carey, 2007; Lee & Sarnecka, 2010; Negen & Sarnecka, 2012; Sarnecka & Gelman, 2004; Sarnecka & Lee, 2009; Wynn, 1990). The following describes the development of number-knower levels in progression:

> **Stage 1: The 'pre-number-knower' level.**
> Infants have yet to distinguish between the meanings of different number words. In Give-N activities, children at this stage will always give one or a handful of things, regardless of the number requested.
>
> **Stage 2: The 'one-knower' level.**
> Children understand that 'one' refers to 1. When asked for one object in a Give-N task, these children will offer exactly one thing, but when requested for any other number, they will give a random quantity between two and the maximum number of objects available.
>
> **Stage 3. The 'two-knower' level.**
> Here, children now understand what 'two' signifies. At this stage, a child offers one object when requested for one' and two objects when asked for 'two', but their responses remain inconsistent when asked for a higher number.
>
> **Stages 4 onwards.**
> The previous stage is repeated at the 'three-knower' level, and finally on the 'four-knower' level. Then, when children grasp the cardinality principle at the 'four-knower' level, they also understand the meaning of all number words in their counting range.

Studies have shown that this progression of number-knower levels is not limited to English-speaking children but has also been observed in Mandarin Chinese-

speaking children (Le Corre, et al., 2003; Li, Le et al., 2003), in Japanese-speaking and Russian-speaking children (Sarnecka, et al., 2007), and in Macedonian-speaking children (Nikoloska, 2009).

Various studies indicate that higher levels, such as 'five-knower', 'seven-knower' or 'eight-knower', do not exist. The explanation for this is proposed to be due to the concept of subitising (see Chapter 2) which states that only the smallest amounts, between one to four items, sounds, flashes of light, etc. can be accurately perceived without counting (Antell & Keating, 1983; Butterworth, 1999; Feigenson, et al., 2002; Feigenson & Carey, 2003, 2005; Scholl & Leslie, 1999; Simon, 1997; Starkey & Cooper, 1980; Wynn, 1992; Uller, et al., 1999). The theory underlying this phenomenon is that humans (and certain other animals) have intrinsic, non-verbal cognitive mechanisms that allow them to precisely represent small amounts up to three or four, however, they have to estimate bigger numbers (e.g., Feigenson et al., 2004). Because these nonverbal systems do not differentiate say between seven and eight items, any numbers higher than four must be determined by counting (Dehaene, 1997).

Nikoloska (2009) pointed out that many children in number-knower studies come from Western countries and were assessed in Western kindergartens, where the curriculum is geared toward school readiness and where counting is often regularly practised. She conducted her own research with kindergarteners in Macedonia, where preschool children are not involved in school preparation activities until the last year of kindergarten, when they are 5 to 6 years old. She also pointed out that, whereas a child's day in a Western European kindergarten is highly planned and academically focused, kindergartens in Macedonia

function in a more *laissez-faire* manner. Children's activities are left open and "[a day] generally passes in free-play, watching cartoons, drawing, physical exercising, learning foreign language or dancing. Around 30 minutes each day children spend learning about various topics, but counting as such is rarely practised" (Nikoloska, 2009, p. 464).

Since counting is an early mathematical skill, Nikoloska (2009) sought to examine the cardinality principle in Macedonian context and whether there was a difference in age and acquired skills compared to Western educated children, that would affect the point at which they grasped the cardinality principle. Her study selected a Give-N task to examine this on the basis that this measure of the cardinality principle is the most commonly used form of testing in the literature, and therefore any age differences identified would be more comparable. The expectation was that the children would go through the same one-, two-, three-, and four-knower phases that all children go through before becoming cardinality-principle-knowers, albeit at a somewhat later age than children who had participated in earlier studies conducted elsewhere. Nikoloska (2009) found that children in Macedonia became cardinality knowers at a later age than the literature reports is the norm, although she did observe five-knowers during the test. These children, when asked for five items, would count, and offer the correct number of items. When asked for six, they would offer a random number of items, but they would not count them. On this basis, Nikoloska (2009) concluded that they were true five-knowers. Children who accurately gave six items were also able to correctly give eight or nine, indicating that they were true cardinality-principle-knowers, as once children learn the cardinality principle, they understand the meaning of all number words in their counting range. Further

research is required to expand on the concept of five-knowers and the relationship between subitising and cardinality principle understanding.

Wynn (1992b) used longitudinal research to evaluate the development of children's comprehension of number words outside of the setting of counting. The study's rationale required children to know the exact meaning of the word 'one', i.e., that it refers to only one object and not to multiple objects. Virtually all of the 2-and-a-half to 3-and-a-half-year-olds passed this pre-test. Successful participants were then given a series of tasks to complete over the course of seven months. To evaluate if they understood that number terms refer to numbers and not to other qualities, these children were shown pairs of images, one representing a single item of a particular sort of object and the other depicting two to six representations of the same item in a different colour. Additionally, all objects shown had the same form in singular and plural, making the number word the only indicator of numerosity. The children were instructed to identify the picture that displayed a specific number of objects while looking at both, for example, they were asked "can you show me the five fish/one fish?". They were also asked questions in which real number words were replaced with nonsensical words, such as "can you show me the zoop fish". These questions elicited random responses.

If children who understand the meaning of one also understand that a number word other than one denotes a specific number (two denotes 2), they should be able to infer that that word does not denote the same number as the word one (two does not denote 1). As such, to answer questions about comparing one object to numerous objects, one might expect that children would use the process of elimination to select the correct picture. For example, a child may be shown a

picture of one fish and a picture of four fish, a number it does not understand yet. The child is then instructed to point out the picture of four fish. If the child understands that four also refers to a number and that different number words refer to different numbers (that they are well distinguished by their respective phonological representation) even if they do not understand the exact meaning of four, they should be able to identify the correct image. If the child does not understand that a specific number word identifies a specific number, they will have no basis for comparing it to one and will reply indiscriminately, just as if the question is posed with a nonsensical word.

To see if they understood the exact number that a particular word denotes, the children were shown a pair of images, one with the number of objects matching the word in question and the other displaying that number plus one. For instance, to see if a child understood the precise meaning of five, they would be given one picture with five fish and another with six fish. The researcher would ask, "can you show me the five fish?" in some trials and "can you show me the six fish" in others. Understanding the concept of one will not help here as, in order to answer correctly, the child has to understand the specific meaning of at least one of the number terms. Almost every time, they correctly pointed out the number asked which demonstrates that they grasped that number words represent numbers and belong to the same semantic class as one. Furthermore, this research indicates that children have an implicit awareness that each number is denoted by precisely one number word. Despite early awareness of this, and in accordance with number-knower development levels mentioned previously, it took nearly a year for children in the trials to learn which words corresponded to which numbers. For example, a 2-year-old who demonstrated that they

understand two (as evidenced by the fact that they never chose the card showing one fish when asked, "can you show me the two fish?") might go for a further year without understanding that 'two' refers to the number 2 and 'three' refers to the number 3. Until that point, when shown a picture of two fish and one of three fish, they would answer the question "can you show me the two fish?" in the same haphazard manner as "can you show me the zoop fish?" (Wynn, 1992b).

4.6. Semantic Singular/Plural Markers

According to Villarroel et al. (2009), the age at which certain number-knower stages are mastered seems related to the different use of semantic singular/plural markers in the language the children in their study grew up with, in this case Basque and/or Spanish. It is noted that both monolingual and bilingual children took part in the study by Villarroel et al., and this may have been a confounding variable. Whilst English calls for the addition of a plural marker in form of a suffix to any description of quantities larger than one, including nouns (apple vs. apples), pronouns (she vs. they), and verbs (is vs. are; does vs. do), and Spanish also adds these markers to adjectives, in Basque, the addition of a plural marker is far from a general rule. In sentences such as: "There are two cars", "You bought three apples", "Give me four pencils", no plural suffixes are used in Basque. Nouns and adjectives stay the same, regardless of whether they reference sets of one or several objects. This feature of Basque is called the indefinite grammatical case[1]. As a result, even when referring to multiple objects, a large percentage of communicational exchanges associated with the indefinite case

[1] Other languages that lack semantic singular/plural distinction include Vietnamese (Lê & Noël, 2020), Japanese and Chinese (Li et al, 2013).

completely lack plural references (Villarroel et al., 2011).

Using the Give-N task, Villarroel et al. (2009) found that, among their sample of 131 children aged 2-3, 70% of the monolingual Spanish speakers had already mastered the two-knower stage or higher, while only 30% of the children who predominantly or only spoke Basque at school and at home had attained that level. In other words, those children whose language had compulsory use of plural markers appeared to enter the number-knower stage in which they understand that one is distinct from two (and above) earlier than children whose language lacks that feature.

4.7. Recap of Number Acquisition

The sections above illustrate how transparency of a number naming system facilitates early learning in young children. The sections also outline how inconsistencies can make counting-word acquisition complicated due to the unique nature of number words. In addition, the cardinality principle has been explored, along with the concept of number-knowers, which refers to the gradual comprehension of the concept that the last number in a counting sequence denotes the total number of items in a set. So far this chapter has defined and described numerical thinking. The remaining chapter deals with the following concepts: dimensionality as it applies to numeration systems, the mental number line, a widely used metaphor for numerical representation in the mind, and how estimation abilities develop.

4.8. Dimensionality

According to Zhang and Norman (1995), one of the essential features of a taxonomy of numeration systems is dimensionality. Number words are linearly connected to numbers in a one-dimensional (1D) system, i.e., each number is represented by a different lexeme. Two-dimensional (1x1D) systems, on the other hand, use a recursive concept to compose number words from a finite set of numerals, such as the base-10 system, which combines base and power/magnitude. The base numbers constitute the main cycle (1–10), which is repeated each time its endpoint (the base) is reached. Most languages specify the number of iterations (the power or multiple of the base) together with the basic numerals to differentiate each cycle (Bender & Beller, 2011). When the quantity to be represented is small and limited, 1D systems will suffice. For example, in body counting systems numbers are physically mapped onto specific body parts in a one-to-one relationship (Lean, 1992; Saxe, 1981; Wassmann & Dasen, 1994). Since the number of body parts is limited, counting among the Yupno in Papua New Guinea, for example, began with the fingers on the left hand, then switched to the right hand, then to the left toes, then the right toes, the ears, the eyes, and so on (Figure 4.1, a), with the counting sequence ending when all body parts were done. However, Bender and Beller (2011) hypothesise that as numbers go up, 1x1D systems become more advantageous to use, because their cyclic patterns help to keep number words compact while drastically decreasing the number of numerals required. Dimensionality also has an impact on mental representation. The bigger the number, the easier it is to express with a 1x1D system.

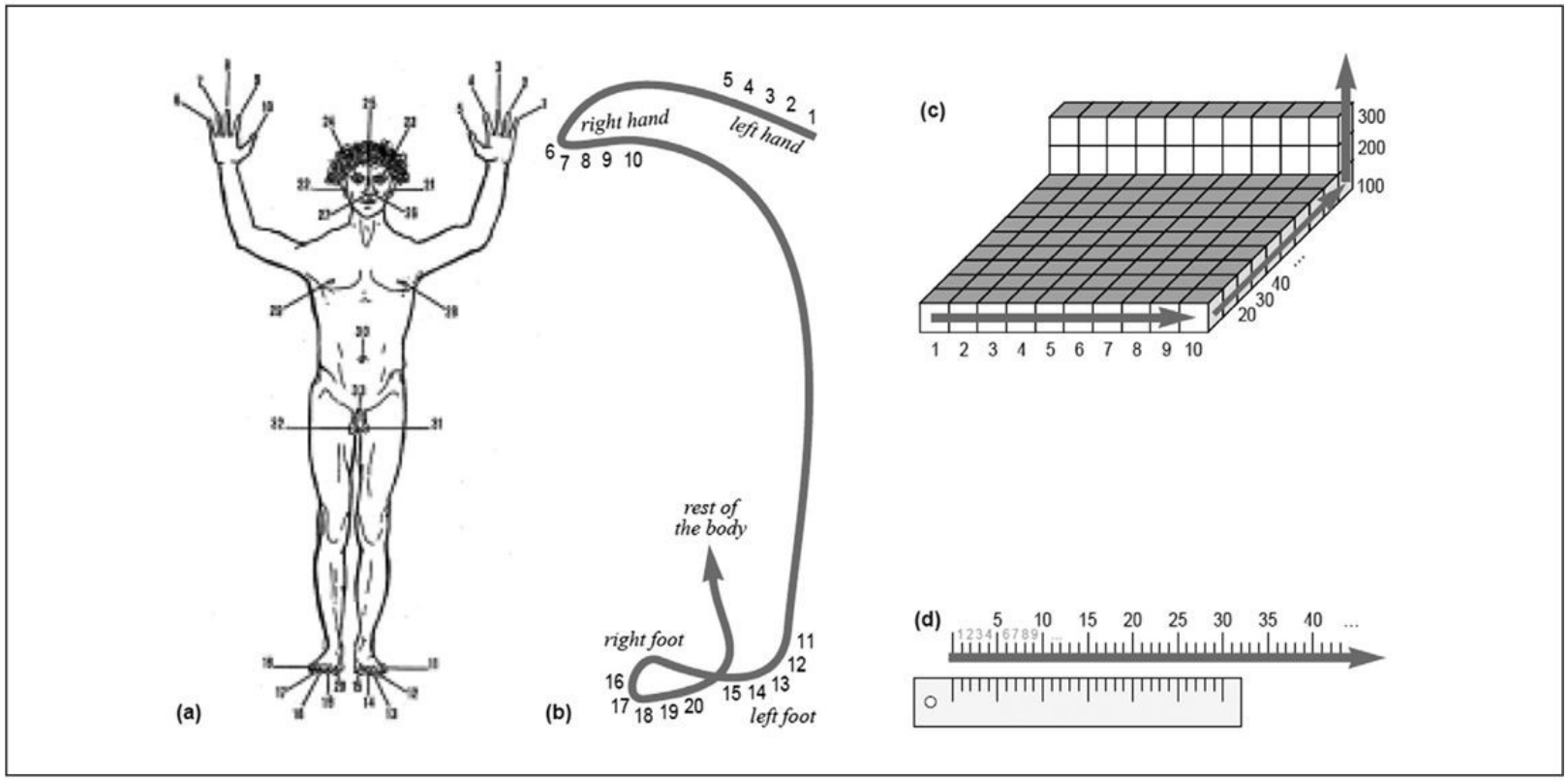


Figure 4.1. Different possible number representations (from Bender & Beller, 2011, p. 13). (a)+(b) Yupno Counting Sequence (Wassmann & Dasen, 1994, p. 84) and number line derived thereof, (c) multidimensional representation inherent in Place-Value-System, (d) straight and linear number line.

A 1D system, like Figure 4.1 (a), maps very well onto a 1D number line like the one represented in Figure 4.1 (b) even if the line is curved slightly because counting proceeds across the front of the body. A 1x1D systems, on the other hand, might be expected to foster two-dimensional representations, like Figure 4.1 (c), which resembles a table with rows and columns.

Representations of 1x1D systems might be imagined as having more than two dimensions since every additional power level adds another dimension. This factor should be even more pronounced in languages with transparent numeration systems like Chinese, and likely becomes less pronounced the less transparent a languages numeration system is. However, there is not yet been enough research on comprehensive number line representation to confirm or reject Bender and Beller's (2011) hypothesis. Nevertheless, most empirical evidence does not appear to support the premise of two-dimensional (or more) representations.

Reynvoet and Brysbaert (1999) discovered that numbers like 10 are equally well primed by 9, their neighbour on the left, and 11, their neighbour on the right. From this, they deduced that single-digit and two-digit Arabic numbers address the same dimensional number line. Dehaene et al. (2008) hypothesised that people's familiarity with measuring instruments such as rulers could be one element nurturing a linear perception. Fixed numerical units are applied to different spatial places along a straight line when measuring (Figure 4.1, d), but, as will be discussed below, the spacing of the mental ruler is not always fixed nor even.

The characteristic that controls the ease with which humans, and certain animals,

differentiate two numbers, is not the absolute numerical distance between numbers, but rather the distance relative to their size. In the early stages of learning, the gap between 8 and 9 is not the same as the distance between 1 and 2. The 'mental ruler' that we use to measure numbers is not graded with evenly spaced markings. It tends to condense bigger numbers into a smaller area. Dehaene (2011) describes how numbers elicit not just a sensation of quantity, but also an inexorable sense of extension in space. The close relationship between numbers and space was evident in Dehaene's number comparison tests (Dehaene et al. 1990) where participants were required to categorise numbers as smaller or larger than 65. To that aim, they used two answer keys, one in the left hand and one in the right. The researchers used different sides for response: half of the subjects answered "larger" with their right hand and "smaller" with their left hand, while the other group answered the other way. This factor had a significant impact: subjects in the "larger-right" group answered faster and made fewer mistakes than those in the "larger-left" group. Subjects pressed the right-hand key quicker than the left-hand key when the target number was greater than 65, and the reverse was true when the target number was less than 65. Large numbers appeared to be automatically linked with the right-hand side of space in the subject's mind and small numbers were automatically associated with the left. The greater the number, the faster right-hand responses are, compared to left-hand replies. Likewise, the lower the number, the greater the tendency to reply more quickly on the left. He termed this phenomenon the 'SNARC' effect, an acronym for "Spatial-Numerical Association of Response Codes". The findings were confirmed in a later study (Dehaene et al., 1993). The absolute magnitude of a number appears to be irrelevant and what matters is how big a number is in comparison to the range of numbers utilised in a given experiment. When only

numbers 0-5 are used, 4 and 5 are preferentially linked to the right. However, 4 and 5 are linked to the left when only numbers from 4 to 9 are used. Additionally, which hand is used to reply is irrelevant, even with individuals with crossed hands and right-sided responses are made with the left hand, the right side of space is linked to bigger numbers. Dehaene's (2011) finding of an inherent link between numbers and space led to the creation of a simple but effective metaphor for the mental representation of numerical quantities: the number line.

The next section looks at mental representations of numbers, specifically the metaphor of the Mental Number Line which is a powerful tool in examining the development of numerical estimation abilities in children. As will be shown, numerical estimation abilities are a good indicator for future mathematical learning. It will be demonstrated in the next section how linguistically transparent counting sequences are an auxiliary to the development of numerical estimation skills and therefore mathematical skills in general.

4.9. The Mental Number Line

According to Dehaene (2011), the Mental Number Line (MNL) is a metaphor whereby numbers are mentally put in a place on a horizontal string. Each place represents a different number, adjacent numbers next to each other in ascending order. Furthermore, the line may be thought of as being oriented in space, with zero at the far left and bigger numbers stretching to the right.

The automatic orientation of numbers in space is matched by the reflexive encoding of Arabic numerals as quantities. The left-to-right orientation raises the question about whether it has its origin in hemispheric specialisation. Do left-

handed people also associate large numbers towards their right side? According to Dehaene's own studies, this does not appear to be the case, performance by left-handed people did not differ from that of right-handed people (Dehaene, 1993, 2011).

Several more studies (Ito & Hatta, 2004; Shaki & Fischer, 2008; Zebian, 2005) have explored the possibility that the origin of the mental orientation of the axis from left to right might depends on cultural variables. For example, Iranian students who had learned to read from right to left as children and lived in France had no preferential association between numbers and direction, as a group. However, individually, their preferred direction of number association depended on exposure to Western culture. Iranian students who had lived in France for five years or more showed the same SNARC effect as native French students, i.e., associating large numbers with the right. In contrast, Iranian students who had been in France for only a few years tended to associate large numbers with the left side of space (Dehaene et al., 1993). The conclusions drawn were: (1) the direction of writing/reading appears to influence the direction of association between numbers and space, and (2) cultural immersion can have major impact on this effect (Dehaene et al., 1993; Ito & Hatta, 2004; Shaki & Fischer, 2008; Zebian, 2005). Using MNL tasks is one tool to examine internal number representation. Number line estimation can be used to assess participants' internal MNL estimations by calculating the discrepancy between the actual location of the target number and the estimate of the participant where participants are required to judge the position of a target number on a blank number line (e.g., Booth & Siegler, 2006; Siegler & Booth, 2004; Siegler et al., 2009). It must be noted, however, that some authors such as Núñez (2008)

dismiss the mental number line as little more than a complex concept mediated by Western culture. Admittedly, research on the specific nature of representations *not* influenced by Western education is still in its initial stages (Bender & Beller, 2011). That being said, using the MNL as a metaphor for internal number representation is deemed appropriate for the purposes of this thesis, which focuses solely on European number, language, and education systems.

4.10. Development of Numerical Estimation

Young children are not very skilled at number estimation. This conclusion can be seen throughout the literature on this topic and has been drawn by researchers studying various properties like 'number of discrete objects' (Hecox & Hagen, 1971), 'distance' (Cohen et al., 1979), 'money' (Sowder & Wheeler, 1989), and 'answers to arithmetic problems' (LeFevreet al., 1993). Various explanations have been proposed as to why children are not skilled at number estimation, one of which may play a large part in this lack of skills: reliance on inappropriate number representation. While it has been hypothesised that children's internal representation of numbers is reflected by their numerical estimation, the nature of number representation varies considerably.

Dehaene (2011) suggests that people of all ages, as well as many animals, rely on a logarithmic-ruler-representation. The difference between logarithmic and linear representations of numbers is that the former exaggerates the distance between the magnitudes of numbers on the low end of the scale while also minimising the distance between the magnitudes of numbers in the middle and upper regions of the scale. This means that, with logarithmic representation, the

psychological distance between the numbers 1 and 75 is greater than the psychological distance between 75 and 1000.

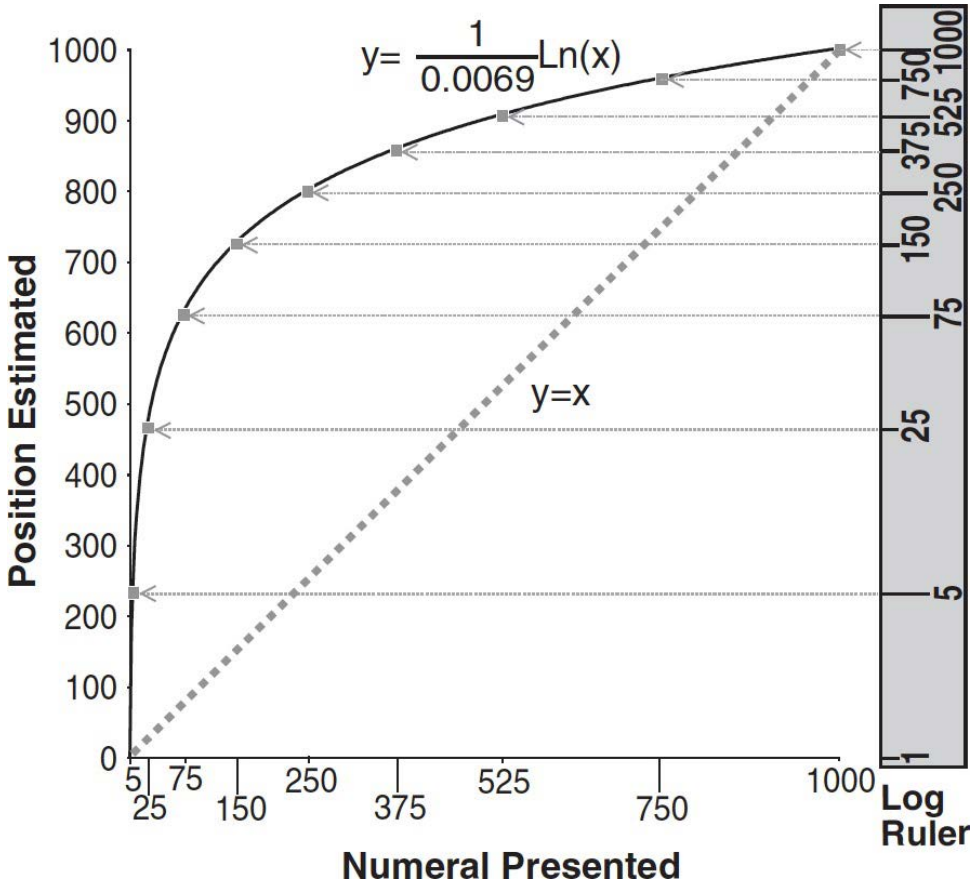
According to Dehaene (2011), the logarithmic-ruler-representation cannot be inhibited as it occurs "as a reflex" (p. 78). For example, it has been observed that when children or adults are asked to draw many lines of any length or to generate numbers at random, both the lines' lengths and the generated numbers, and even the time it takes to solve magnitude comparison problems, fit a logarithmic function (ibid.).

Gibbon and Church (1981) proposed the accumulator model, which Brannon et al. (2001) hypothesised to be used by children and adults of all ages. The accumulator model claims that numbers, as well as other quantities, are represented as equally spaced and linearly increasing magnitudes with scalar variability. Scalar variability involves representations becoming noisier, and thus more variable, with increasing magnitude. That is to say, the amount of variability increases linearly together with the numbers being represented. In 2001, Huntley-Fenner showed that the means and variances on a dot-estimation task done by 5- to 7-year-olds also fit the prediction of this model.

Another hypothesis was proposed by Case et al. (1996) who suggested that children of different ages use different representations, but children of a given age use only one kind of representation. For example, children of 4 and 5 years of age would only possess a qualitative central conceptual structure for representing numbers (e.g., "there are a few toys in this box and in that one there are many toys"), whereas children aged 6 and older would consistently rely on a representation with a linear structure. Case and Okamoto (1996) found that 4-

year-olds generally cannot accurately estimate which of two single-digit numbers is closer to a third, whereas 6-year-olds generally are able to do so. This structure was depicted by Case and Sowder (1990) as a number line with linearly increasing magnitudes, and they posited that accurate numerical estimation is impossible before it is acquired. Siegler and Booth (2004) postulated that while all three of the aforementioned models provide plausible accounts, it is very unlikely that any of them alone describes how humans represent numbers and other quantities. Rather, it appears more probable that humans know and make use of multiple ways to represent numbers, and the type of representation used is influenced by situational and contextual variables (Figure4.2). In this multiple representation hypothesis, the types of representations that children rely on in any given situation are thought to change with age and numerical experience. This hypothesis follows in the footsteps of classical developmental theorists like Vygotsky (1934/1962), Piaget (1954), and Werner (1957), all of whom postulated that development is characterised by patterns of parallel changes across different ages and variable time spans.

Figure 4.2. Predicted estimates of the logarithmic and linear ruler models, when both are constrained to pass through the endpoints of the number line (Siegler & Booth, 2004, p. 429).



Siegler and Booth (2004) suggested that children might be able to use logarithmic and accumulator representation from infancy. As children gain more experience with counting, arithmetic, and other numerical contexts in the formal number system, they may add to their representation repertoire more linear representations and a variety of other representations of number categories, such as odd/even, square/non-square, decade-names/other. In other words, Siegler and Booth (2004) thought that there is not just one representation of numbers, but

that humans must possess multiple and varied representations, coexisting and competing. Starting off with logarithmic representation, with time, age and experience, children learn to rely on the most appropriate representation according to situational needs.

Using multiple numerical magnitude representations would be most advantageous under different circumstances. Logarithmic representations are most useful at an early stage because, in an unfamiliar range of numbers, they discriminate numerical magnitudes on the low end of the range to a greater extent. When first learning the numbers 0-100, children encounter the numbers at the low end of this range more often in situations like counting exercises, single-digit addition and subtraction problems, and other early numerical activities. In these circumstances, it is useful to have enhanced discrimination at the end of the range where these low numerical magnitudes are located (Ashcraft & Christy, 1995; Hamann & Ashcraft, 1986). In comparison, linear representations discriminate the magnitudes of the larger numbers in the middle and at the high-end range more distinctly. The distinct discrimination becomes more important when children start to grapple with multi-digit addition and subtraction, as well as single-digit multiplication and division problems where they are more likely to encounter numbers of greater magnitude more frequently (Siegler & Booth, 2004).

Siegler and Opfer (2003) provided support for the hypothesis of multiple representations. They asked second, fourth, and six-graders as well as adults to do number line estimation tasks. The number lines had endpoints marked with either 0 and 100 or 0 and 1000 but were otherwise unmarked. The researchers found a developmental change in children between the second and sixth grade

on the 0–1000 lines in that reliance shifted from logarithmic to linear representation. For second graders, 95% of the variance in median estimates could be accounted for by the logarithmic representation, while the best fitting linear representation only accounted for 63% of the variance. For sixth grades and adults, in contrast, the best linear representation was a perfect fit and accounted for 100% of the variance in median estimates for both age groups. The logarithmic representation, on the other hand, only accounted for 73% and 78%, respectively.

That being said, when using the 0–100 lines, the researchers also found evidence that the logarithmic representation was not the only representation of which second graders could make use. For 43% of this age group, the variance of the median estimates fit the *logarithmic* function when using the 0–1000 number line, and the *linear* function when using the 0–100 line.

According to Siegler and Booth (2004) "estimation patterns, and presumably the representations used in generating them, vary with age, context, and task" (p. 431). In their study they asked kindergarteners, first, and second graders to complete line estimation tasks on 0–100 lines. The children also completed a Stanford Achievement Test (SAT-9) for mathematics three to four months after the experimental session. They found that almost all kindergarteners generated estimates in logarithmic patterns, almost all second graders generated estimates in linear patterns, and that first graders split evenly between both patterns of estimation. Furthermore, controlling for age and grade, they also found that individual differences in number line estimation skills were strongly related to individual differences in maths achievement, especially for first and second graders (Siegler & Booth, 2004). Siegler et al. (2009) confirmed these

conclusions in their study when they consistently observed a developmental shift, or "stable unit of cognitive change" (p.149), between 5 and 8 years of age. At that age, children increasingly shift from logarithmic representation to rely on a linear representation of numerical magnitudes more and more. This progression is age dependant and occurs first for smaller number ranges before linear representation of larger number ranges emerges (Booth & Siegler, 2008). Children between kindergarten and second grade progress from logarithmic to linear representation in the 0–100 range, and between second and fourth grade in the 0–1000 range. Changes can occur over a period of years, and sometimes within a single experimental session, which indicates, according to Siegler et al. (2009), that children show more or less abrupt changes in the progression from one representation of numerical magnitudes to the other, and that these changes can encompass a large range of numbers. Representations of very large numbers, on the other hand, remain largely compressed on a logarithmic representation, even in adults (Banks & Coleman, 1981; Dehaene & Marques, 2002).

The remainder of this chapter examines the educational implications of numerical mental representation on the MNL and it will be demonstrated how linguistically transparent counting sequences are an auxiliary to the development of numerical estimation skills and therefore mathematical skills in general. The following section addresses representation on the mental number line and linguistic effects on number line representation.

4.10.1. Representation on the Mental Number Line

Research suggests that number representation revolves around the MNL, and that children's MNL representations correlate with and are causally linked to

arithmetic performance (e.g., Link et al., 2014; Booth & Siegler, 2006, 2008; Siegler & Booth, 2004). Okamoto and Case (1996) proposed that the development of linear ruler representation at around the age of six allowed children to solve a wider range of mathematical problems than they were able to solve before.

This, in turn, would mean that individual differences in MNL estimation skills in 5–8-year-olds are linked to individual differences in maths achievement scores. Siegler and Booth's (2004) study also showed that individual differences in number line estimation skills were strongly related to individual differences in maths achievement in kindergarteners, first, and second graders. The accuracy of number line estimation is also associated with some early problems in number learning. (Muldoon et al., 2011). The use of suitable numerical magnitude representation appears to be crucial for a variety of educationally relevant outcomes. The accuracy of children's magnitude comparisons, and therefore their capacity to learn, for example, solutions to new addition problems, is significantly correlated with the linearity of their MNL estimations with whole numbers, and thus accuracy for larger numbers (Booth & Siegler, 2008). Additionally, several studies found a connection between accuracy of MNL estimation and overall scores on mathematics achievement tests, (Booth & Siegler, 2006, 2008; Siegler & Booth, 2004).

In assessing participants' internal MNL estimations using number line estimation, the discrepancy between the actual location of the target number and the participants' estimate is calculated. During *non-verbal number line estimation*, the target number is presented visually instead of verbally. This is particularly advantageous to assess linguistic effects on MNL representation without the

confounding factor of verbal comprehension, which might be separately influenced by language (Helmreich et al., 2011). An irregular counting system such as English or German includes number words that do not show a one-to-one correspondence with the Arabic numerals, and therefore need to be learned separately. The degree to which the spoken and written number systems of languages match onto the Arabic numerals, are one explanation for results in international comparisons of children's arithmetical performance (Dowker et al., 2008; Dowker & Roberts, 2015; Muldoon et al., 2011; Siegler & Muldoon, 2008). A regular counting system, such as the Chinese system, might facilitate children's learning of higher numbers as well as increase their understanding of the place value system (Muldoon et al., 2011). Though systematic research on the subject began relatively recently, the idea that counting systems might influence arithmetical performance was proposed by Edgeworth and Edgeworth as far back as 1798. They posited that the irregularities of the English language's counting system may be a disadvantage when learning arithmetical skills (Edgeworth & Edgeworth, 1798).

The fact that some countries with transparent counting systems show superior results in mathematical performance does not indicate a causal relationship between the two. Other cross-cultural differences, such attitudes to mathematics or academic skill in general, methods of and/or the amount of time dedicated to mathematical education in school, and even the economic or political situation in a country, could be contributing factors to these differences in performance (Dowker et al., 2008; Dowker & Roberts, 2015; Geary et al. 1996; Laski & Yu, 2014).

In their 2014 study, Laski and Yu investigated the relative importance of

language and education to the development of numerical knowledge. Their participants were monolingual Chinese kindergarteners and second graders (educated in China, dialect not further specified but living in Shanghai) and bilingual Chinese-American kindergarteners and second graders (educated in America, fluent in both Chinese – dialect not further specified – and English). All were given a number line estimation and a mental addition task. Overall, both groups' number line estimation abilities were found to be one to two years more advanced than previously tested American children's number line estimations (Siegler & Mu, 2008), but the children educated in China performed better than the ones educated in America. The researchers drew the conclusion that education may have a greater influence on numerical development than the linguistic structure of counting system. However, while the Chinese-American children spoke Chinese at home and were fluent in Chinese, they had received the entirety of their mathematical education and were tested during this study in English, whereas the Chinese children had been educated in and were tested in Chinese. This difference alone obscures whether the findings were due to education rather than linguistic factors. Muldoon et al. (2011) also found a difference between Chinese and Scottish 4- to 5-year-old participants, but stated that Chinese children's number estimations are not more accurate than that of *older* Western samples of equivalent ability .However, when they included short number lines such as 0-10 and 0-20, they found that the Scottish children performed better than the Chinese children.

There are few studies available that explore the relationship between language transparency and MNL estimation in children, all of which are described above and below in this chapter. Most of these studies include Chinese as one of the

languages and some do not explicitly state whether they are comparing monolinguals with bilinguals. There have been no studies exploring the relationship between language transparency and MNL estimation in children that have been conducted explicitly on multilinguals. The following are the only two studies available where neither of the languages is Chinese. The second study, Helmreich et al. (2011), is one of the studies discussed in detail later as part of this scoping review.

4.10.2. Linguistic Effects on Number Line Representation

In order to bypass the problem of different cultures and educational systems completely, Dowker et al. (2008) and Dowker and Roberts (2015) looked at pupils in Wales attending either Welsh-or English-medium schools. All of the children lived in the same area and in similar cultural environments, and all studied maths according to the National Curriculum of England, Wales, and Northern Ireland, "[to make it] possible to study the specific effects of language on arithmetic, independently of other educational and cultural differences" (Dowker & Roberts, 2015, p. 2). There are few studies available using a language other than Chinese as the transparent example, and Dowker and Roberts (2015) is the only one using Welsh. The Welsh language contains two counting systems, an irregular one and a regular one, and only the latter is used to teach arithmetic in schools (Roberts, 2000). This counting system is as transparent the Chinese one, for example, 11 is *un deg un*, which literally translates to one-ten-one and 22 is *dau degg dau* or two-tens-two. Dowker et al. (2008) tested 6-and 8-year-old Welsh children with different degrees of exposure to the Welsh language, i.e. Welsh at home and school, Welsh at home only, English at home only. The children were tested with a variety of number processing tasks, and it was found

that both groups of Welsh-speaking children performed better at comparing two-digit numbers. However, there was no difference between the groups in terms of arithmetic test performance. This led the researchers to the conclusion that linguistic effects do not have a global effect on arithmetic performance, but rather only affect specific aspects of it. Dowker and Roberts (2015) used transcoding and non-verbal number line estimation in order to investigate whether the internal representation of numbers in children is affected by the transparency of the counting system. The participants were Welsh-English bilingual Year 2 pupils, 20 of whom attended an English-medium school and 20 a Welsh-medium school. They found no difference in overall arithmetic abilities between the two groups, but the pupils who were taught in Welsh were better at non-verbal line estimation. This was most apparent in number estimations over 20, and this cohort had a smaller range of errors. The researchers link these results to the transparency of Welsh counting system.

Helmreich et al. (2011) looked at two different languages entirely, Italian which has a counting system that is mostly transparent (but not to the degree of Chinese and Welsh), and also German. In German, the potentially most confusing property is inversion, akin to English - teen numbers (e.g., 19 = “nine” (9) -“teen” (1)). However, in German, all tens and units up to 99 are inverted. For example, 21 is *ein-und-zwanzig*, which literally translates to one-and-twenty and 56 is *sechs-und-fünfzig* or six-and-fifty. The participants in Helmreich et al.'s (2011) study were Italian-and German-speaking first graders, educated in Italy and Austria, respectively. The two groups of children had an age difference, with the Austrian children being older, with a mean age of 7 years 4 months, whereas the mean age of the Italian children was 6 years 11 months. Austrian children

often enter first grade one year later than Italian children, however, despite this, the school curriculum regarding mathematics is virtually identical in both countries. In this study they used a paper-pencil version of a *number to position task* for the number line task, the children were required to indicate their estimate of a given number with a pencil on an empty number line. The line was 10cm in length, the left end was labelled 0 and the right end was labelled 100. The researchers were interested in the influence of language properties on non-verbal number line estimation. More specifically, they were interested in whether the inversion property of most German multi-digit number-words might have any influence on estimation ability. To ensure that no verbal number word was involved in the task at any level, presentation or response, the numbers to be estimated were printed in Arabic notation in the middle of the page above the line. Before the start of the estimation task, children were told not to count or use any other strategy except to estimate the position of the presented number on the line. The number 50 was used in a trial run, but no feedback or correction was given at any point. Results showed two language-specific differences: (1) the estimates of the Italian children were more accurate overall, despite the children being significantly younger, and (2) the Italian children performed particularly better when inversion errors led to large estimation errors in German children as Italian did not make errors like guessing 48 when the target number was 84, but the Austrian children did frequently. The researchers concluded, that even in non-verbal settings, children's mental number line organization is indeed influenced by language properties.

4.11. Conclusion

This chapter explored how the transparency of a number naming system aids early learning in young children and how inconsistencies can impede the acquisition of counting words. It also presented a discussion of the cardinality principle and the concept of number-knowers, or the gradual realisation that the final number in a counting sequence represents the total number of items in a set. In the second part of the chapter, the idea of dimensionality as it relates to numeration systems was introduced. The widely utilised metaphor of the mental number line was employed to demonstrate how numbers are mentally represented. Finally, the chapter addressed two aspects pertaining the importance of the mental representation of numbers for numerical education: how representation on the mental number line correlates with and is causally linked to arithmetic performance and linguistic effects such as transparency and inversion on number line representation. The next chapter explores the framework of differentiated instruction (DI).

Chapter 5: Differentiated Instruction

5.1. Introduction

This chapter introduces the framework of differentiated instruction (DI) and how it may enable teachers and other educational practitioners to meet the needs of different learners and help them succeed in mathematical learning, especially in classrooms with diverse aptitudes and in multilingual classrooms. In cases where the language of instruction differs from the first languages of learners, students are mentally taxed with learning new mathematical concepts while also learning the language in which these new concepts are taught. In these cases, differentiated instruction can support both teachers and learners.

5.2. Differentiated Instruction

Differentiated instruction (DI) has developed over the past few decades as a central framework for meeting the needs of diverse learners in contemporary classrooms (Tomlinson, 2014) It is premised on the recognition that students vary in their readiness, prior knowledge, cultural background, and language proficiency, and that instruction should be adapted accordingly. The growing emphasis on inclusive education and equity has brought DI to the forefront of pedagogical research and practice (Tomlinson, 2001; 2014). Rather than prescribing a uniform method, DI encourages teachers to design learning experiences that provide multiple entry points, different levels of challenge, and varied modes of engagement and expression. The following subsections examine, the rationale behind DI and its theoretical underpinnings as well as

critiques and challenges, how DI can be implemented for mathematical learning, how it can be applied to multilingual classrooms, and which strategies are advantageous in such environments.

5.2.1. Rationale for Differentiated Instruction

The modern classroom is characterized by heterogeneity (Essien, 2016; Roberts & Olarte, 2025). Students differ in cognitive abilities, learning styles, motivation, cultural identities, and linguistic repertoires. In multilingual contexts, students may be simultaneously developing proficiency in the language of instruction and in subject-specific academic registers (Gibbons, 2002). In mathematics and science classrooms, learners enter with varied experiences and background knowledge, requiring carefully designed supports.

Traditional models of teaching often assume that students are a homogeneous group and rely on whole-class instruction targeted at an "average" student. DI, by contrast, rejects this assumption. It recognizes that teaching at a single pace or level risks leaving some students behind while failing to sufficiently challenge others (Hall, 2002). Tomlinson (2001; 2014) emphasized that effective differentiation involves proactive planning, where teachers anticipate learner diversity and prepare multiple pathways for engagement. The rationale for DI therefore rests not only on inclusivity but also on effectiveness. Differentiated approaches can foster higher student engagement, improved motivation, and deeper learning outcomes when compared with rigid instructional designs (Smale-Jacobse et al., 2019). Carol Ann Tomlinson (2001; 2014) organizes DI around four interconnected elements: content; process, product, and learning environment.

These four elements provide a useful means of distilling the core aspects of DI:

1. Content refers to adjusting what students learn. For example, advanced students may explore more demanding topics, while others continue to focus on core concepts.
2. Process is related to how students engage with material. This can include collaborative work, independent projects, or use of technology.
3. Product accounts for how students demonstrate understanding of subject matter, such as through essays, presentations, or creative projects.
4. Learning environment relates to creating a classroom climate that is flexible, supportive, and respectful of diversity.

The theoretical foundations of differentiated instruction are closely linked to the sociocultural perspectives of Lev Vygotsky and the constructivist contributions of Jerome Bruner. Vygotsky's concept of the zone of proximal development (ZPD) (1978) and Bruner's work on scaffolding (1976) provide essential theoretical grounding for understanding why and how differentiated instruction may be apt to meet the needs of different learners in various contexts. Both Vygotsky's theory of the zone of proximal development (1978) and Bruner's scaffolding theory (1976) provide compelling theoretical rationale for tailoring learning experiences to students' diverse needs. Tomlinson's framework translates these theoretical underpinnings into practical elements related to content, process, product, and environment, while maintaining a focus on equity and responsiveness. In order to examine the theoretical groundings of differentiated instruction, the following subsections explore these theoretical foundations, links them to contemporary understandings of DI, and discusses key principles, critiques, and challenges.

5.2.2. Theoretical Underpinnings

According to Vygotsky (1978), knowledge is not simply transmitted from teacher to student but co-constructed through interaction. Social interaction

provides the context for cognitive development: learners engage with more knowledgeable peers or adults who guide them toward higher levels of functioning. Vygotsky's (1978) sociocultural theory therefore positions learning as a fundamentally social process. Central to this perspective is the concept of the zone of proximal development (ZPD), defined as the distance between what a learner can achieve independently and what they can achieve with assistance from a more knowledgeable other (adult, teacher, more experienced peer). Instruction is most effective, Vygotsky argued, when it targets this zone. If tasks are too easy, little learning occurs; if they are too difficult, frustration and disengagement result. Within the ZPD, learners are challenged but supported, enabling them to internalize new skills and concepts. In Vygotsky's theory of learning and cognitive development, the ZPD is proposed as a means if supporting the child to transition from spontaneous or everyday concepts (ones which are rooted in a child's direct everyday experiences, unsystematic) to scientific or formal/academic concepts (learned through formal schooling and systematic instruction).

The ZPD has profound implications for classroom instruction (Mahn, 2015), including language education (Lantolf & Poehner, 2014), clinical learning (Kantar, Ezzeddine, & Rizk, 2020) and mathematics (Norton & D'Ambrosio, 2008; Scott, 2016). Together the body of research on the ZPD suggests that teachers in a range of contexts must continually assess students' current levels of understanding and adjust the level of challenge accordingly. It is this ongoing adaptation to learners' levels of understanding and readiness for further challenges that forms the core of DI. Teachers who differentiate are, in essence, identifying the ZPD for each learner or group of learners and structuring

instruction so that students are supported in moving beyond what they can do unaided. Lantolf, Panahi, and Mohebbi (2024) assert that much of Vygotsky's research indicated processes of development emerging during very short periods of time (Poehner & Infante, 2016), and that in order to deliver more effective education, grouping them according to the size of their ZPD is required (Lantolf, Panahi, and Mohebbi, 2024, p.265).

Building on Vygotsky's work (1978), Jerome Bruner and colleagues introduced the metaphor of scaffolding (Wood, Bruner, & Ross, 1976) to describe the temporary support provided by teachers or peers that enables learners to accomplish a task they could not complete alone. The support is gradually reduced as competence increases, much like the scaffolding used in construction is removed once a building can stand independently. Scaffolding takes many forms: modelling, questioning, feedback, prompts, or breaking tasks into manageable steps. In mathematics, a teacher might scaffold by demonstrating a problem-solving method, asking guiding questions, and gradually shifting responsibility to the student. In language-rich subjects, scaffolding may involve providing sentence frames, visual aids, or structured group discussions. Scaffolding ensures that tasks are adapted to learners' needs and learning supports are dynamic, adjusting as competence grows. Within a differentiated classroom, scaffolding allows teachers to provide varying levels of assistance simultaneously, ensuring that students at different points in their development all remain engaged.

It is important to note here that research and practitioner publications sometimes conflate Vygotsky's concept of the Zone of Proximal Development (1978) with metaphor of scaffolding proposed by Bruner, Wood and Ross (1976). While

both highlight the importance of guided support in learning, they stem from distinct theoretical origins and serve different purposes. The ZPD remains a developmental construct emphasizing social interaction as the engine of cognitive growth. In contrast, scaffolding was proposed as a metaphor for the temporary, adaptive support an expert provides to help a novice complete a specific task. Over time, educational literature in particular has tended to treat scaffolding as the pedagogical realization of the ZPD.

However, theorists such as Smagorinsky (2018), Chaiklin (2003) and Daniels (2001) argue that this conflation misunderstands and simplifies Vygotsky's dynamic, culturally mediated framework into a procedural teaching technique. In their critiques, the ZPD is not a step-by-step instructional method, but a means of conceptualizing how learning potential develops through socially shared activity. Lantolf and Thorne (2006) further emphasize that scaffolding represents only one form of assisted performance within the broader sociocultural perspective. Critics also contend that aligning scaffolding too closely with the ZPD can misrepresent Vygotsky's view of development, which includes internalization and peer collaboration, not solely adult-led guidance (see for example: van der Veer & Valsiner, 1991; Little, 2004; Phillips, 2014).

Tomlinson's (2001; 2014) four interconnected elements of DI (content; process, product, and learning environment) links with Vygotsky's theoretical work and Bruner's proposal around support in the classroom. Targeting instruction within the ZPD ensures that content is pitched appropriately according to learners' needs, especially regarding development of scientific concepts. Scaffolding represents adjustments to the process of learning. The products or outcomes of learning may vary depending on students' linguistic or cognitive strengths, while

the learning environment reflects the social nature of development emphasized in Vygotskyan sociocultural theory (Tomlinson, 2001; 2014). As the defining feature of DI is responsiveness, where teachers continually monitor student progress and adjust instruction accordingly, formative and diagnostic assessment methods allow for an exploration of learner's conceptual development. Conducting formative and diagnostic assessment enables teachers to identify readiness, interests, and learning profiles (e.g. flexible grouping to allow students to work with peers at similar readiness levels for certain tasks, while at other times heterogeneous grouping to encourage peer scaffolding). The theoretical grounding of this kind of pedagogy exemplifies Vygotsky's emphasis on what a learner can achieve with support, not simply on what they can do independently. In this sense, DI operationalizes sociocultural theory by embedding assessment within instructional decision-making.

Differentiated instruction also aligns with other theories of learning. For example, constructivist perspectives (e.g. Fosnot, 2013). emphasize that learners actively build knowledge rather than passively receive it. DI supports this by providing varied experiences and opportunities for students to engage with concepts in meaningful ways, a tenet that can be traced back to Piaget (1970). Humanistic theories, such as those of Maslow (1965) and Rogers (1957), also highlight the importance of addressing students' individual needs and fostering self-actualization. DI reflects these principles by valuing learners' interests and profiles. Moreover, contemporary research on Universal Design for Learning (UDL) (Navaitienė & Stasiūnaitienė, 2021; Levey, 2023) parallels DI by advocating for multiple means of representation, engagement, and expression (Meyer, Rose, & Gordon, 2014). While UDL is often framed in the context of

accessibility and special education, it shares with DI the goal of creating flexible learning environments responsive to diversity.

5.2.3. Critiques and Challenges

While DI is theoretically compelling, its implementation raises challenges. Critics argue that differentiation demands extensive preparation and constant adaptation, which may be unrealistic in classrooms with large numbers of students or limited resources (Marshall, 2016). Smale-Jacobse et al. (2019) note that empirical evidence for DI is mixed, with some studies reporting significant benefits of DI and others showing that it has limited impact. Suprayogi et al.'s (2017) study connected teachers' actual implementation of DI to a complex array of factors: their self-efficacy regarding DI, teaching beliefs, professional experience, professional development activities, certification status, and classroom size. As per the results of the regression analysis, 39% of the variation in DI implementation was significantly associated with self-efficacy beliefs regarding DI, as well as with larger classroom sizes and stronger constructivist beliefs. Another critique concerns equity. While DI aims to promote inclusion, poorly designed differentiation may inadvertently lower expectations for struggling learners. For example, consistently assigning simplified tasks without providing opportunities for growth risks reinforcing achievement gaps (Boaler, 2016). In her response to a case study by Koomen (2016) about a boy with learning disabilities in an inclusive seventh grade life science classroom, Bannister (2016) claimed to make the disparities of DI apparent. She outlines how this method limited the boy's learning and involvement by categorizing him as less competent and of lower status and contends that this approach encourages the reproduction of societal status hierarchies within the classroom.

Effective DI must therefore balance support with challenge, ensuring that all students have access to high-level thinking and conceptual development. Teachers' beliefs also influence implementation. Some educators may perceive DI as additional workload or as conflicting with standardized curricula and assessments. In 2020, Maulana et al. conducted a cross-cultural study to gain empirical understanding differentiated instruction and other related aspects of teaching quality. Their research involved observing teachers' classroom practices in two high-performing educational systems, the Netherlands and South Korea. The study found that DI can be regarded as a separate area of teaching quality in various national contexts. However, in the observed countries, teaching quality posed its greatest challenge. Professional development is therefore crucial, not only for equipping teachers with strategies but also for fostering a mindset that values responsiveness and equity. It is clear that the challenge of implementing DI successfully is related to translating the theory into practice.

Teachers often ask what differentiation looks like in concrete terms and how abstract principles can be applied to specific subjects. Mathematics, with its unique blend of abstraction, hierarchy, and cultural positioning, offers a useful testing ground for exploring how differentiation works in real classrooms. The next section examines differentiated instruction in mathematics, showing how the theoretical ideas discussed here take shape in day-to-day teaching.

5.3. Differentiated Instruction and Mathematical Learning

Mathematics is a subject that is often described as abstract, sequential, and cumulative and students' experiences of learning mathematics vary widely. Some children arrive at school with strong foundational skills and high

confidence, while others may struggle with basic concepts, feel anxious, or lack opportunities for engagement (Paras, 2001, Rameli & Kosnin, 2016). In heterogeneous classrooms, this range of readiness levels presents a challenge for teachers (Jourdain & Sharma, 2016).

Differentiated instruction offers practical strategies to ensure that all students, regardless of their starting point, can participate in meaningful mathematical learning. This section explores how DI operates in mathematics classrooms, with a focus on the practical application of Vygotsky's zone of proximal development (1978) and Bruner's metaphor of scaffolding. Classroom strategies, case studies, and equity considerations, with an emphasis on supporting diverse learners in concrete ways, are also highlighted.

Mathematics learning does not proceed uniformly as students bring varied prior experiences, cultural perspectives, and problem-solving strategies. Research shows that differences are not only cognitive but also motivational and affective (Kim et al., 2014; Hanin & Van Nieuwenhoven, 2016; Carmichael et al., 2017; Abín, et al. 2020; Zhang et al., 2024). Some students may excel in procedural fluency (carrying out mathematical procedures) yet struggle with conceptual understanding, while other students may demonstrate creativity in problem-solving but lack confidence in formal representations (mathematical symbols, expressions, representations such as algebra etc.) (Boaler, 2016). It is also worthwhile to look beyond language 'errors' to identify the underlying reasoning in order to differentiate successfully. Moschkovich (2011) describes an instance which serves as a prime illustration of this. Alicia, a middle school student and English learner whose first language was Spanish, was assigned to describe a pattern relating the dimensions of a rectangle to its perimeter. While

she struggled to find the correct term in both English and Spanish and eventually invented word *rángulo*, she nonetheless successfully communicated a sophisticated mathematical relationship. By using gestures to trace long rectangles and referring to concrete drawings, she demonstrated how a rectangle's perimeter expands as its length (with a fixed area) increases. Alicia's understanding would seem inadequate if the focus was merely put on improving her vocabulary. However, it is clear that Alicia is engaging in a high-level mathematical discourse, namely describing covariation, by diversifying the evaluation process to incorporate non-verbal resources like gestures and objects (Moschkovich, 2011).

This diversity has important implications, as whole-class instruction aimed at a so-called "average or standard learner" risks leaving many pupils behind. DI responds to this diversity by offering multiple routes into mathematical thinking, ensuring that students engage at a level appropriate to their needs. Muthomi and Mbunge's (2014) explored the effectiveness of DI for students' mathematics achievement and whether students' achievements differed when they were taught using the DI approach. They utilised the quasi-experimental Solomon Four-Group design and mathematical achievement tests to collect data from 374 Kenyan secondary school students at eight different schools. The results showed that DI had a significant positive impact on students' maths achievement.

The central question then becomes how to design tasks and supports that support every learner. For example, a student who has mastered multiplication but not division is ready for scaffolded instruction that bridges these concepts. Instruction targeted below this level (e.g., repeated practice of addition) would be

unproductive, while instruction too far ahead (e.g., algebraic equations) would be inaccessible. Teachers can address learners' growth in the ZPD in mathematics classrooms through:

- Pre-assessments and diagnostic tasks to identify students' current knowledge. (Tomlinson, 2014)
- Tiered assignments that align with varying levels of readiness. (Tomlinson, 2001)
- Guided practice where teachers or peers provide structured support. (Wood, Bruner, & Ross, 1976)
- Dynamic grouping that allows learners to move in and out of groups depending on the concept at hand. (Hall, 2002).

The dynamic and individualised nature of the ZPD means that differentiation must be ongoing. Research on the notion of scaffolding as effective theoretical underpinning of effective mathematics pedagogy was shown by Smit (2013) who found that scaffolding in mathematics not only aids conceptual understanding but also supports academic language use. She asked the research question: How can teachers in multilingual primary classrooms scaffold pupils' language required for mathematical learning? To this end she looked into if the implementation of strategies aimed at supporting the development of subject-specific language in a multilingual upper primary mathematics classroom resulted in whole-class scaffolding that could be recognized by its main features, considering the long-term aspect. Her empirical analyses indicated that this was indeed the case: The teacher diagnosed and performed strategies responsively over time, and she significantly transitioned to linguistic independence. When teachers structured discussions carefully and modelled mathematical vocabulary, students engaged more deeply in both content and communication. Effective scaffolding in mathematics involves more than providing hints; it is also about structuring tasks and dialogue so that learners gradually develop independent reasoning. Examples include:

- Worked examples followed by partial problems reduce cognitive load and help learners build problem-solving schemas by providing structure before expecting independent application (Sweller et al., 2011; Renkl, 2014).

- Manipulatives and visual models offer concrete ways for students to grasp abstract mathematical concepts, particularly in early learning and multilingual settings (Smit, 2013). Their effectiveness increases when paired with guided instruction (Carbonneau et al., 2013), although teachers must ensure students connect these tools to symbolic understanding (Uttal, 1997).

- Guiding questions support metacognition and allow students to reflect on their problem-solving processes. They are an essential tool in developing reasoning and understanding in mathematics classrooms (Boaler & Brodie, 2004). Strategic questioning also facilitates meaningful classroom discussions (Chapin et al., 2009).

- Mathematical talk allows students to clarify and extend their thinking, promoting both language acquisition and deeper understanding – especially for multilingual learners (Smit, 2013; Moschkovich, 2010). Sfard (2001) also argues that learning mathematics is a discursive process, where knowledge is constructed through communication and participation in mathematical practices.

Several of these strategies can be utilised in when implementing DI in mathematics classrooms. Firstly, tiered assignments allow students to work on problems of varying complexity. For instance, in a geometry unit, all students may explore properties of triangles, but advanced learners might prove theorems while others focus on identifying relationships through measurement. The teacher ensures that each tier is equally engaging and linked to core objectives (Hall, 2002). Secondly, open-ended tasks provide multiple entry points and solution paths. For example, asking students to design as many rectangles as possible with an area of 24 units allows learners to use multiplication, division, or factorization depending on their readiness (Boaler, 2016). Open tasks encourage creativity and allow differentiation without labelling students. (Suurtamm et al., 2010; Boaler, 2016). Using grouping strategies are a third means of achieving learning goals. As mentioned above, homogeneous groups

may be useful for targeted support while heterogeneous groups can promote peer scaffolding, where more advanced learners explain concepts to others. Teachers can rotate groupings to avoid fixed hierarchies (Smale-Jacobse et al., 2019). Fourth, differentiated questioning, where teachers pose layered questions that vary in complexity, can be employed. For example, when exploring linear functions, one student may be asked to identify slope from a graph, while another may be challenged to interpret slope in a real-world context. Differentiated questioning ensures that all students participate at an appropriate level. (Wood, Bruner, & Ross, 1976; Tomlinson, 2014).

Although DI in mathematics is a useful way to ensure that teaching is tailored to learners' levels of understanding and readiness, there are some challenges associated with its implementation. Boaler (2016) cautions against practices that inadvertently track students into "low" or "high" groups, reinforcing inequities. Differentiation should never mean lowering expectations, rather, the goal should be to maintain a high-level of cognitive demand while providing different pathways to success. One way to address this challenge is to use problems accessible to all but extendable to advanced level learners. For example, exploring patterns in a number sequence can begin with simple arithmetic but extend to algebraic generalizations (Boaler, 2016). This ensures that differentiation does not separate students into rigid ability groups. Another consideration is the cultural relevance of mathematics problems. Research suggests that students engage more deeply when problems connect with their lived experiences (Gutstein, 2006). Differentiated instruction should therefore not only address cognitive readiness but also incorporate diverse cultural perspectives into mathematical tasks. Teachers often express enthusiasm for the idea of DI but find

its implementation challenging. Mathematics teachers, in particular, cite time constraints, curricular pressures, and large class sizes as barriers (Suurtamm et al., 2010). Professional development is essential to equip teachers with practical tools, strategies for flexible grouping, and methods for designing open-ended tasks.

Differentiated instruction in mathematics illustrates how sociocultural theory and its understanding of conceptual development can be operationalized in classrooms through tiered tasks, open-ended problems, flexible grouping, and scaffolding. These strategies ensure that learners encounter mathematical ideas at an appropriate level of challenge and support. Yet many mathematics classrooms can also be linguistically diverse, and for multilingual learners the demands of understanding mathematical content are inseparable from the demands of navigating academic language. Addressing this intersection requires teachers to think beyond differentiation for cognitive readiness alone. They must also attend to the ways in which language scaffolds mathematical reasoning and participation. Building on the insights from mathematics instruction, the next section outlines how differentiated instruction can be extended to multilingual classrooms, drawing on the contributions of scholars who emphasize the integration of language and content support.

5.4. Differentiated Instruction in Multilingual Classrooms

The increasing linguistic diversity of classrooms worldwide has reshaped the educational landscape. Migration, globalization, and shifting demographics mean that many students now learn academic subjects in a language that is not their first. Mathematics, science, and literacy instruction all become more

complex when students must simultaneously acquire subject knowledge and the language of instruction. According to the UNESCO (2016) Global Education Monitoring Report Policy Paper, entitled "If You Don't Understand, How Can You Learn?", "[c]hildren should be taught in a language they understand, yet as much as 40% of the global population does not have access to education in a language they speak or understand."(p.1). In mathematics teaching or learning, activities such as reading, writing, listening, and discussing are central to understanding. For every educator, language serves as a means for presenting and communicating concepts and procedures, for reading and resolving texts and issues, and for evaluating mathematics performance.

Research findings that focus specifically on language competencies in mathematics education illuminate the central role of language in this context. (Pimm & Keynes, 1994). A systematic review conducted by Setati et al. (2009) on multilingual classrooms in South Africa identified only a small number of publications which focused on multilingual classrooms. On the other hand, an abundance of studies had been conducted regarding language issues in the teaching and learning of mathematics, particularly at the high school level. However, studies examining the function of language in mathematics during the early grades (Grades 1–4) are limited in comparison to those focusing on this aspect for older learner cohorts. Essien's (2018) systematic review on research carried out in Kenya, Malawi, and South Africa from 2006 to 2016 on the topic of language in early grade (Grades 1–4) mathematics teaching and learning showed that longitudinal studies examining the effect of language on mathematics teaching and learning are scarce, as are investigations into how teachers in the early grades should be trained to teach mathematics. Roberts and

Olarte (2023) consider the gaps in teacher education in their study which described the need for preservice teachers in secondary education to be prepared more intently to integrate language and content learning in their mathematics classrooms, given the high numbers of multilingual learners.

In multilingual contexts, the language of the classroom is itself a barrier or resource that must be addressed. Teachers are called not only to support conceptual learning but also to support language development, ensuring that students can access, engage with, and express academic ideas. This section explores the practical application of DI in multilingual classrooms, highlighting strategies, case studies, and implications for teacher education and policy, with literature reviewed at the intersection of language and content instruction. Multilingual classrooms are not a new phenomenon, but they have become more common and visible in recent decades. In urban centres, schools may have dozens of home languages represented.

Some children arrive with strong literacy skills in their first language, while others have interrupted schooling or limited literacy in any language. Teachers face the task of teaching a curriculum designed with monolingual assumptions to groups of students with diverse linguistic repertoires (de Jong, 2011).

According to de Jong (2011), the quality and content of teacher preparation for specialist language teachers (English as a Second Language, bilingual teachers) of English Language Learners (ELLs) teachers must be examined. Her study raises questions about the acceptance of mainstream classrooms as English-only, monolingual learning environments in the context of preparing teachers for mainstream education. Even though English predominates as the instructional

medium in mainstream classrooms, a monolingual approach poses challenges because it overlooks the importance of ELLs' native language resources in the learning process. Multilingual learners may also experience marginalization when their languages and cultures are devalued. Instruction that assumes native-like proficiency in the language of instruction risks excluding them from participation in academic discourse. Equity therefore requires an approach that treats linguistic diversity as an asset rather than a deficit (Yi, Son & Chang 2025). Roberts and Olarte (2023) describe now multilingual learners have experienced "historic patterns of exclusion, such as receiving less cognitively demanding work" (p. 467).

Turkan, de Oliveira, Lee, and Phelps (2014) introduced the concept of disciplinary linguistic knowledge (DLK), which refers to the specialized understanding teachers need about the language demands of their subject. For example, mathematics requires precision in vocabulary, syntax for expressing relationships, and discourse structures for argumentation.

Teachers who lack awareness of these demands may inadvertently obscure content for multilingual learners. Turkan and Liu (2012) examined multilingual students' performance on inquiry-based science assessments and found that language barriers, rather than content misunderstandings, explained many of the difficulties that the learners experienced. This finding underscores the need for scaffolding that integrates language and content. Rather than simplifying content or lowering expectations, teachers must provide differentiated supports that allow learners to engage with rigorous material while developing academic language proficiency.

5.5. Differentiation Strategies for Multilingual Classrooms

Earlier in this chapter the strategies for implementing DI in mathematics classrooms were outlined. This subsection describes the implementation of DI in multilingual settings, combining scaffolding for content and language. To guide this support, practitioners can adopt Moschkovich's (2007) framework that views bilingual learners through three distinct lenses:

- Acquiring vocabulary: This narrow view sees mathematical communication as primarily about learning words and translates word problems as a prerequisite for conceptual work.

- Constructing multiple meanings: This view recognizes that students must navigate between every day and mathematical registers (e.g., "prime" in everyday life vs. "prime number" in mathematics).

- Participating in discourse: This sociocultural lens views communication as involving more than just words; it includes gestures, objects, and social practices.

Moschkovich (2007) states that by shifting to a discourse perspective, teachers are able to uncover the resources students already bring to the classroom, such as code-switching or home-language metaphors, rather than focusing on what they lack. The following strategies extend the principles described above in mathematics to encompass linguistic diversity.

Firstly, language scaffolds are structured supports that help learners access and use academic language. Examples include:

- Sentence starters (e.g., “The graph shows that…”), which guide students in framing responses.
- Word banks and glossaries, which ensure access to subject-specific vocabulary.
- Visual aids, including diagrams, charts, and graphic organizers, which reduce reliance on text alone.
- Modelling of academic discourse, where teachers explicitly demonstrate how to explain reasoning or construct arguments.

These scaffolds align with Bruner’s body of work (1976), which suggests that students learn best when there is a gradual transfer from heavily reliance on provided structures, to students over time internalising these structures to use them independently. Gibbons (2002) highlighted the role of structured classroom talk in supporting multilingual learners. By providing sentence frames, modelling interactions, and setting clear roles in group work, teachers created opportunities for learners to participate in academic discourse: “In short, an integrated program takes a functional approach to language by placing its teaching focus on language as the medium of learning rather than on language as something separate from content” (Gibbons, 2002, p 208). Differentiation, in this example, involved adapting the process of learning rather than the content, ensuring equitable participation. Turkan and Liu (2012) showed that multilingual learners struggled with inquiry-based science assessments, not because of a lack of scientific understanding, but because of linguistic complexity. They investigated the effect of students' English language status on their performance in inquiry science. In their study, differential item functioning (DIF) analysis was performed concerning English Language Learner (ELL) status on an inquiry-based science assessment. In total, 1,396 students from seventh and eighth grades took the science test, of which 313 were ELLs. The results indicated that, by and large, non-ELLs performed significantly better

than ELLs. Of the four items that exhibited DIF, three were advantageous to non-ELLs and one was advantageous to ELLs. The component that benefited ELLs included a graphic depiction of a scientific idea in a family context. The researchers concluded that constructed-response items can assist ELLs in expressing scientific reasoning in their own words. When creating assessments for ELLs and offering them instruction, developers of assessments and educators should be mindful of the potential interplay between language-related difficulties and scientific subject matter. Differentiated strategies such as glossaries, visual supports, and collaborative scaffolding can help learners to demonstrate their knowledge more effectively.

Secondly, translanguaging is increasingly examined in the literature with reference to refers to the use of multiple languages within a learning context (Gandara & Randall, 2019). Their research investigated a translingual approach to administering a mathematics assessment in the Democratic Republic of Congo. They argue that it can take multilingual students several years to reach the same level of language proficiency as their monolingual peers and therefore assessing them in a language they do not fully grasp is inappropriate. Multilingual students do not act like several monolinguals, so translated tests are not an adequate solution. It is important for developers of tests to create evaluations that facilitate the use of multilingual students' complete linguistic repertoires and their natural linguistic practices. The results of their study indicate that a translingual administration is more suitable than a traditional one to evaluate mathematics.

García and Wei (2014) argue that students draw on their full linguistic repertoires, rather than compartmentalized "separate" languages. In practice, this might

involve allowing students to brainstorm in their home language before producing work in the language of instruction or using bilingual word walls and resources. Translanguaging validates students' identities and lowers cognitive load, enabling deeper engagement with content (García & Wei, 2014).

Thirdly, as touched on earlier, grouping strategies can be a key means of effectively implementing DI. In multilingual classrooms, heterogeneous groups can be powerful because peers scaffold each other's language use. However, homogeneous groups may be helpful for targeted language support. Teachers can vary groupings depending on the task, ensuring that no student is permanently assigned to a "low" group. Fourthly, task adaptation can be employed to modify linguistic demands without diluting learning content. For instance, in an exercise, all students might carry out the same procedures, but multilingual learners could receive structured prompts, while others complete the exercises in a more open-ended format. The cognitive challenge remains the same and only the language scaffolding differs. Finally, differentiation in multilingual classrooms can also incorporate students' cultural knowledge. Culturally relevant tasks connect learning to students' lived experiences, enhancing engagement and affirming identity (Gay, 2010). For example, in mathematics, teachers might design word problems that reference familiar contexts drawn from students' communities.

Differentiated instruction provides a framework for making multilingual classrooms more equitable. By integrating language and content supports, teachers can ensure that all students – regardless of their linguistic background – engage meaningfully with academic subjects. Research on disciplinary linguistic knowledge (Turkan et al., 2014) highlights the need for teachers to understand

not only content but also the language through which content is constructed and expressed. The strategies discussed above, including language scaffolds, translanguaging, flexible grouping, task adaptation, and culturally relevant pedagogy, demonstrate how differentiation can be applied concretely in multilingual contexts. The findings above demonstrate that when teachers scaffold both content and language, learners participate more fully in classroom discourse and achieve deeper understanding.

5.6. Conclusion

The abstract concepts taught in mathematics present inherent challenges for many learners. When learners must acquire a new language on top of learning new mathematical concepts, and the language of instruction is not their first language, the challenges are more complex. In such situations, DI can provide valuable support for both teachers and learners. To tackle this intersection, teachers need to consider more than just differentiation based on cognitive readiness. They should also consider how language supports mathematical reasoning and involvement. Language can also pose a challenge for small children learning the language specific names given to numbers in their first language. The different quirks and irregularities of number naming could also increase the mental load of learners, and when this is not mitigated early in the learning journey, it could result in a ripple effect for further mathematical learning.

Chapter 6: Research Design

6.1. Introduction

This brief chapter presents an overview of research design utilised in this study and provides details of conceptual and procedural reasoning that connects the thesis to the methods, data, and analytical strategies used in both reviews. Relevant literature and guidance related to scoping reviews and systematic reviews is outlined, after which, the rationale for the research design chosen for this study is provided. The following chapters include the methodology employed for each review.

6.2. Research Design Considerations

Research design must account for the complexity, contextuality, and interdisciplinarity of the phenomena under investigation (Creswell & Creswell, 2018, Denzin & Lincoln, 2018). Questions related to language, learning, and cognition often span multiple theoretical traditions, methodological approaches, and educational contexts. As a result, a carefully structured and transparent research design is essential to ensure coherence, rigour, and cumulative knowledge-building.

In recent decades, evidence synthesis has become an increasingly important methodological approach within the social sciences. Rather than generating new primary data, review-based designs systematically collect, evaluate, and synthesise existing research in order to clarify patterns, identify gaps, and inform theory, policy, and practice. Within applied linguistics, where research is frequently dispersed across subfields and methodologies, structured review

approaches play a crucial role in consolidating knowledge and guiding future inquiry (Macaro, 2019). Evidence synthesis encompasses a family of related designs, each serving distinct but complementary purposes. Systematic reviews aim to comprehensively identify, appraise, and synthesise research evidence according to transparent and replicable procedures. They are especially well suited to addressing focused research questions, evaluating the strength of evidence, and drawing conclusions about effectiveness or impact. However, systematic reviews presuppose a degree of conceptual clarity and definitional stability within the field. When terminology is inconsistent, theoretical frameworks are diverse, or the scope of existing research is unclear, a systematic review alone may be premature or overly restrictive.

For these reasons, the present thesis adopts a sequential review design, beginning with a scoping review and followed by a systematic review. Scoping reviews are particularly appropriate in emerging or conceptually fragmented areas of research. Their primary purpose is to map the extent, range, and nature of research activity; clarify key concepts and definitions; examine how research has been conducted; and identify gaps in the literature. Unlike systematic reviews, scoping reviews do not typically focus on assessing the quality of evidence or estimating effect sizes. Instead, they provide a structured overview of the landscape of scholarship.

In the context of this thesis, the scoping review serves an exploratory and clarificatory function. Given the interdisciplinary nature of research at the intersection of language, learning, and mathematics education, the initial step is to establish how core constructs have been defined, operationalised, and investigated. This mapping process allows for a more precise delimitation of the

research problem and informs the development of focused research questions for the subsequent stage.

Building on the insights generated in the scoping phase, the second stage comprises a systematic review designed to address a more narrowly defined set of questions. Whereas the scoping review identifies patterns and gaps across the field, the systematic review evaluates and synthesises evidence in relation to specific criteria. This sequential design ensures both breadth and depth: the scoping review establishes conceptual and methodological orientation, and the systematic review provides analytical rigor and evaluative synthesis. By structuring the research design in this way, the thesis aligns exploratory mapping with systematic evaluation. The sequential approach enhances transparency, reduces the risk of premature narrowing of the evidence base, and strengthens the overall coherence of the study. Together, the two phases contribute to a comprehensive and methodologically robust account of number names and numerical development in a multilingual context.

6.3. Systematic Review Research Family

There are various review types within the systematic review family of methods, and each type of review has distinct purposes, strengths and weaknesses. Choosing a review type for research depends on several factors, like the research topic, the research questions, the sort of evidence that is available, the level of rigour that is necessary, and the time and resources available for the study. High-level synthesis of evidence is best suited to the systematic review framework and meta-analyses (Arksey & O'Malley, 2005; Grant & Booth, 2009; Levac et al., 2010; Peters et al., 2020) whilst investigating the state of the

art of novel, intricate, or theoretical topics is best suited to scoping, narrative, and critical reviews as they provide greater flexibility.

6.3.1. Systematic Reviews

Systematic reviews answer address specific research questions by identifying, appraising, and synthesising all relevant studies on a topic. They have a narrow scope, focussing on specific questions, and often they are conducted using predefined inclusion criteria. The desired result of a systematic review is to end up with a comprehensive and structured review of all relevant studies related to a specific research question and to this end, the methodology involves a predefined protocol, an exhaustive and systematic search strategy, and a rigorous appraisal of study quality, often limited to peer-reviewed studies. It also often includes a meta-analysis to be able to statistically combine results from multiple studies. The strengths of a systematic review are that it provides a high level of evidence due to its rigorous methodology (Guyatt et al., 1995), reduces the risk of bias by only incorporating reliable sources through its systematic search and selection processes (Higgins & Green, 2008; Macaro, 2019; McDonald, 2020), and is suitable to inform and facilitate decision-making in clinical and policy contexts by summarising the best available evidence (Cook et al., 1997). Its weaknesses, on the other hand, are that its rigorous process is time-consuming and resource-intensive (Ahmad, 2024; Tsertsvadze et al., 2015). It may be too narrowly focused on a specific question, potentially missing broader contextual information (Garritty et al., 2024; Mohseni et al., 2022), and its overall quality depends heavily on the availability of high-quality primary studies (Ahmad, 2024; Shaheen et al., 2023).

6.3.2. Rapid Reviews

Systematic reviews can also be conducted in the form of a rapid review to provide a quick synthesis of evidence to inform decision-making under time constraints (Moons et al., 2021). A rapid review is a streamlined version of a systematic review that uses abbreviated methods (such as limiting the search to specific years or databases) to quickly gather evidence on a topic (Ganann et al., 2010; Moons et al., 2021). Alternatively, a review of a set of systematic reviews can be completed, known as an umbrella review (or overview of reviews). This design is intended to summarize and synthesize evidence from multiple systematic reviews on a broad topic (Belbasis et al., 2022). Umbrella reviews focus on higher-level evidence and provide an overarching synthesis across multiple questions, which is often used to inform high-level guidelines and policy recommendations (Choi & Kang, 2023; Fernandez et al., 2025).

6.3.3. Meta-Analysis

The purpose of a meta-analysis is to statistically combine results from multiple studies to provide a pooled estimate of effect sizes (Borenstein et al., 2021). It has a rather narrow scope which is focused on quantitative studies with similar methodologies in order to allow for statistical aggregation. Typically, it is part of a systematic review, but it can be a stand-alone method if sufficient studies for analysis exist (Gurevitch et al., 2018). It uses quantitative methods and a search strategy similar to that of systematic reviews, but requires that studies have used comparable methods and outcome measures. The strength of the meta-analysis lies in increasing statistical power by combining data from multiple studies and providing a precise estimate of the effect size; therefore, it is a helpful research design for resolving existing inconsistencies in the literature (Higgins & Green,

2008). It can also be used to identify patterns, trends, or moderators across studies. Conversely, one of its weaknesses is that it requires studies included in the analysis to be sufficiently homogeneous in terms of methods and outcomes, which can be very limiting (Huedo-Medina et al., 2006). Meta-analyses may also be affected by publication biases, where only positive results are published and included (Rothstein et al., 2005). This design also requires a high level of statistical proficiency, as complexity in methods can lead to misinterpretation if not executed properly. Conducting a meta-analysis was an unsuitable research design for this thesis given the small amount of published peer-reviewed studies in this field, which is as of yet a rather new and emerging area of linguistics.

6.3.4. Scoping Reviews

Scoping review research design, on the other hand, allows researchers to map the breadth of research on a broad topic, and to identify key concepts, types of evidence, and research gaps by exploring a wide range of studies and sources without focusing on a specific outcome or question (Arksey & O'Malley, 2005). The design and search strategies are overall less stringent than with systematic reviews, as scoping reviews are often used to clarify working definitions and conceptual boundaries, and there is usually no formal quality assessment of included studies (Tricco et al., 2018). The methodology is flexible and iterative; as there are no strict guidelines for study quality assessment, the focus is on breadth rather than depth of the given topic (Levac et al., 2010). They provide an exploratory entry point into an area of research where gaps can be charted and future research agendas informed.

The search strategy used in scoping reviews is comprehensive and may include diverse sources, including grey literature (Peters et al., 2020). The strengths of scoping reviews are it is an apt means of exploring new or emerging research areas, flexible in terms of methodology and research question refinement, and well-suited to identify gaps in knowledge and therefore inform future research directions (Arksey & O'Malley, 2005; Munn et al., 2018). The weaknesses, however, are that since there is no assessment of the evidence's quality, their usefulness for decision-making is limited, their broad scope may result in less detailed analyses, and overall, the findings tend to be more descriptive than analytical or conclusive (Tricco et al., 2018).

Scoping reviews, therefore, facilitate a thorough mapping of the main ideas, categories of evidence, and gaps in the current body of knowledge, in contrast to systematic reviews, which concentrate on providing answers to particular research topics (Munn et al., 2018). In areas where research is still in its early stages, this method is very helpful since it allows researchers to spot patterns, define terms clearly, and provide the groundwork for future, more targeted studies.

6.4. Rationale for Sequential Design

In order to better meet specific goal or this research study, a scoping review was chosen as the design of the first phase of the study in order to establish the state of the art in the literature. Scoping reviews are unique within the systematic review family because of their adaptability, wide range, and emphasis on outlining study areas rather than providing answers to particular problems, as outlined in the description of various review types above. They are especially

helpful in determining research gaps, examining novel subjects, and providing guidance for upcoming investigations, which aligns with the subject matter and the aim of this thesis. While scoping reviews are not as thorough and rigorous as systematic reviews, and therefore cannot be used to draw firm conclusions or offer advice, they can be carried out within a variety of frameworks which offer organised methods to guarantee rigour, transparency, and repeatability.

The systematic review research design consists of a rigorous, transparent, and reproducible methodology for identifying, evaluating, and synthesising research evidence (Page et al., 2021). This approach is now fundamental to evidence-based practice across diverse fields, including applied linguistics and the social sciences (Snyder, 2019). As a research design, systematic reviews address focused research questions frequently framed under terms like PICO or SPIDER (Garritty et al., 2024). The PICO framework was used in this study, as a means of identifying Populations, Interventions, Comparators and Outcomes (see Chapter 8).

The search strategy in these reviews must be extensive and repeatable to mitigate risks like publication bias (Ahmad, 2024). Standardized screening, typically involving two independent reviewers, ensures methodological rigour (Page et al., 2021). Furthermore, modern frameworks emphasize that the quality of the review depends on the assessment of 'risk of bias' in primary studies, ensuring that decisions are founded on evidence reliability (Shaheen et al., 2023).

In the first phase of this research study, a scoping review was conducted to ascertain the extent of the published literature on the potential effects of number-naming conventions on number acquisition and numerical abilities in young

children. Conducting the scoping review in this initial phase helped to identify pertinent age groups, educational contexts, and outcome measures associated with differences in mathematics learning. However, because the measurement instruments used across the reviewed studies varied considerably, the strength of claims that could be drawn from Phase 1 was limited. This variability highlighted the need for further investigation through a more focused systematic review of the literature.

The scoping review also facilitated the identification of relevant theoretical frameworks that may explain the efficacy of differentiated instruction (DI) techniques, including constructivist learning theories, language-mediated number processing, and cognitive load theory. Finally, findings from the scoping review informed the aim and research question of the systematic review undertaken in Phase 2 of the study.

The systematic review conducted in Phase 2 enabled an evaluation of the effectiveness of differentiated instruction in specific contexts and for specific learner groups. It was informed by the results of the scoping review, which indicated that DI is often supported on the basis of learner language and cognitive diversity. Following the scoping review's focus on number-naming conventions and number acquisition, undertaking a systematic review of differentiated instruction in mathematics was both necessary and methodologically appropriate. This systematic review examined DI as a potential instructional response within the identified problem area, conceptualised through scoping review findings that highlighted learner heterogeneity driven by linguistic and cognitive characteristics. Together, the two phases of the study established a coherent research trajectory, beginning with a broad conceptual

mapping of the field and culminating in an evidence-based systematic review with both theoretical and practical contributions.

6.5. Conclusion

This chapter provided an overview of the research designs utilised in the study, outlining both the scoping review design and the systematic review design and locating them within the overall systematic review family of methods. As scoping review methodology does not permit evaluative findings regarding particular instructional tactics, a systematic review was incorporated as a means of moving the focus from a broad mapping of issues to a more thorough, outcome-oriented assessment of Differentiated Instruction as a pedagogical approach in the context of number naming conventions and linguistic diversity. Chapter 7 and Chapter 8 presents each review type, research questions, search strategies, and findings.

Chapter 7: Scoping Review (Phase 1)

7.1. Introduction

This chapter starts with the methodology used to conduct this scoping review. This is followed by a detailed summary, preparing the way for an analysis of the dataset. For ease of reference and reading, the eight studies identified in this scoping review will be henceforth referred by number as shown below in Table 7.3. The second part of the chapter will present the general characteristics, similarities and differences in the dataset of eight articles, through a process described as charting the data. Based on the data extracted from the eight articles, the aim here is to map the main directions of the evidence and to identify key themes, findings and possibly also gaps.

7.2. Arksey and O'Malley Methodological Framework for Scoping Reviews

The Arksey and O'Malley Framework (2005) is perhaps the most widely cited and foundational framework for conducting scoping reviews. It is one of the seminal frameworks for conducting scoping reviews and laid the groundwork for how scoping reviews are conceptualised and conducted. Many of the more recent frameworks build on or refine this initial work. It is designed to clarify and map the key concepts in a research area and identify gaps in the existing literature. It outlines the following steps for researchers:

1. Identifying the research question: Clearly define what you are exploring.
2. Identifying relevant studies: Determine which studies are relevant to your question.
3. Study selection: Develop criteria for including or excluding studies.
4. Charting the data: Systematically extract and chart data from the studies.
5. Collating, summarising, and reporting the results: Analyse and present

the findings.
6. Optional consultation exercise: Consulting with stakeholders to refine the findings.

Levac, Colquhoun, and O'Brien (2010) build on the Arksey and O'Malley (2005) framework, offering refinements to increase rigour. They emphasise the importance of a clear rationale for each methodological choice and suggests a more iterative approach to refining the research question and study selection criteria, recommending a more detailed and systematic approach to data synthesis and reporting. Similarly, the Joanna Briggs Institute (JBI) Methodology for Scoping Reviews (2015) provides detailed guidance and checklists for conducting scoping reviews, focusing on transparency and reproducibility. Their recommended steps are:

1. Defining the review objective/question: Establish the purpose and objectives.
2. Developing a protocol: Create a clear protocol before conducting the review.
3. Eligibility criteria: Set criteria for including/excluding studies.
4. Study identification: Conduct a comprehensive literature search.
5. Data extraction: Systematically extract data.
6. Data presentation: Present data in a narrative and/or tabular format.
7. Discussion of results: Interpret the results in relation to the review objective.

Peters et al. (2020) further refine the JBI (2015) methodology, providing an updated framework for scoping reviews. They emphasise the iterative nature of the scoping review process, encourages the inclusion of a broader range of sources beyond peer-reviewed articles, such as grey literature, and highlight the importance of stakeholder involvement. PRISMA-ScR (Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews) (Tricco et al., 2018) is an extension of the PRISMA guidelines. PRISMA-ScR provides a checklist for reporting scoping reviews which

includes essential items and optional items that guide authors in reporting scoping reviews. It focuses on transparency in reporting to improve the reproducibility and quality of scoping reviews. This checklist was used to format the presentation of data and findings that follow.

The main framework utilised for this scoping review is that of Arksey and O'Malley (2005), since it offers an adaptable, all-encompassing method that is especially helpful for exploratory research. However, compared to other frameworks, it is less explicit in some methodological elements such as data synthesis, even if it outlines key stages. For this reason, the PRISMA-ScR (Tricco et al., 2018) checklist was extremely helpful, as it provided thorough guidelines on decision-making and reporting guidelines which raised the standard and transparency of this scoping reviews. It was designed with a broad exploratory focus rather than a concentration on a specific grade level. Consequently, the selection framework did not prioritise a particular grade level from the outset. The search strategy for the scoping review (Phase 1) was designed to be intentionally broad and iterative, using the Arksey and O'Malley (2005) framework to map the full breadth of evidence in a field that is still in its infancy. This comprehensive mapping allowed to produce significant learner variability, by giving students different conceptual starting points and linguistic structures like transparency and singular/plural distinctions.

7.3. Identifying the Research Question

The methodological framework followed in this scoping review recommends starting with wide definitions for the study population, in order to support a wide search strategy. The research question posed in the scoping review was:

What does the literature indicate about relationship between number-naming conventions in different languages and number acquisition in young children?

7.4. Search Strategy

Step two of Arksey and O'Malley's methodological framework involves identifying relevant sources. This section elaborates on the process of searching and selecting the sources, specifically:

- Information sources
- Search Strings
- Inclusion and Exclusion Criteria
- Data Extraction

7.5. Information Sources

To find relevant sources for this scoping review, the following databases were utilised to search for peer-reviewed research published in English. These databases were accessed through library subscriptions at Trinity College Dublin.

- ProQuest: Linguistics and Language Behavior Abstracts (LLBA)
- Web of Science Core Collection (using exclusion settings for the following categories: Computer Science Artificial Intelligence, Computer Science Interdisciplinary Applications, Computer Science Theory Methods, Engineering Electrical Electronic, Acoustics, Computer Science Information Systems, Rehabilitation, Audiology Speech Language Pathology, Sociology or Literature, Literature Romance or Anthropology, Environmental Sciences, Physiology, and Psychology)
- Via EBSCOhost, the following databases were accessed:
  - Academic Search Complete
  - British Education Index
  - Education Full Text
  - ERIC
  - General Science Full Text
  - Humanities Full Text
  - MLA Directory of Periodicals
  - MLA International Bibliography

- OmniFile Full Text Mega
- APA PsycArticles
- APA PsycInfo
- Readers' Guide Full Text Mega
- Social Sciences Full Text
- Google Scholar for supplementary, 'grey literature' search

7.6. Search Strings

Finding studies for a scoping review involved a number of methodical, repeatable search terms that are inserted into the above databases listed above. To execute the searches, the following four search strings were entered into the database search engines in both the titles and abstracts of publications, using the OR function. These search strings were ultimately combined using the AND function. The asterisk (*) included in these search strings is used to broaden the scope of the search. It functions as a placeholder and enables the search for different iterations of words starting with the same letter(s), e.g. the search term *convention** will return *convention*, *convention-s*, *convention-al*, *convention-ally*, etc.

String 1

Convention* OR Custom* OR Tradition* OR Manner* OR Style* OR Practice* OR Mode* OR Usage* OR characteristic* OR structure* OR system* OR Use*

String 2

Language* OR Multilingual* OR Bilingual* OR "linguistic difference*" + language acquisition AND SLA

String 3

Acquisition* OR Learn* OR acquire* OR knowledge OR understand* OR competence*

String 4

“number*-name*” OR “number* name*” OR “number representation*” OR “number-representation*” OR “number concept” OR “number-concept” OR represent* OR “number word*” OR “number-word*” OR “numer* system*” OR “numer*-system*” OR ”counting system*” OR ”counting-system*” OR “number system” OR “number-system” OR “number naming system*” OR “number-naming system*” OR “number naming-system*” OR “number-naming-system*” OR Number* OR numerical* OR numeration* OR numeral* OR numeric* OR count*

For the search conducted in the “Web of science” database, the search term “Child*” was added to eliminate the particularly numerous results involving adult participants. To reduce the number of initial results, the following fifth search string was added: “Relationship* OR Link* OR Tie* OR Relation* OR Association* OR Interconnection* OR Dependenc* OR Connection* OR Correlation*”. This attempt was unsuccessful.

7.7. Inclusion and Exclusion Criteria

The study's inclusion and exclusion criteria are presented in Table 7.1. Unlike a systematic review, a scoping review's inclusion and exclusion criteria are developed in parallel with the search process and are guided by the researchers' increasing familiarity with the literature available.

Table 7.1. Inclusion and exclusion criteria

Inclusion Criteria	Exclusion Criteria

• Original research published in:
  o peer reviewed books
  o conference papers & proceedings
  o dissertations & theses
  o scholarly journals
• Published in English
• Focus on the factors affecting early numeracy development in children
• Focus children learning number systems
• Studies that looked at the relationship between language and number system learning
• Any date of publication
	• Research conducted among child participants with autism, neurodiversity, and/or language impairment
	• Research conducted with deaf children and/or children learning and using sign language
	• Research that included or focused on adult learners/participants,
	• Research that focused exclusively on arithmetic/maths development
	• Studies focused on transcoding between languages.
	• Involved machine learning or artificial intelligence

7.8. Study Selection

The third step in the methodological framework is to use the inclusion and exclusion criteria developed above and identify relevant studies for analysis. The PRISMA diagram displayed in Figure7.1 displays the article selection process to identify suitable articles for inclusion in the scoping review

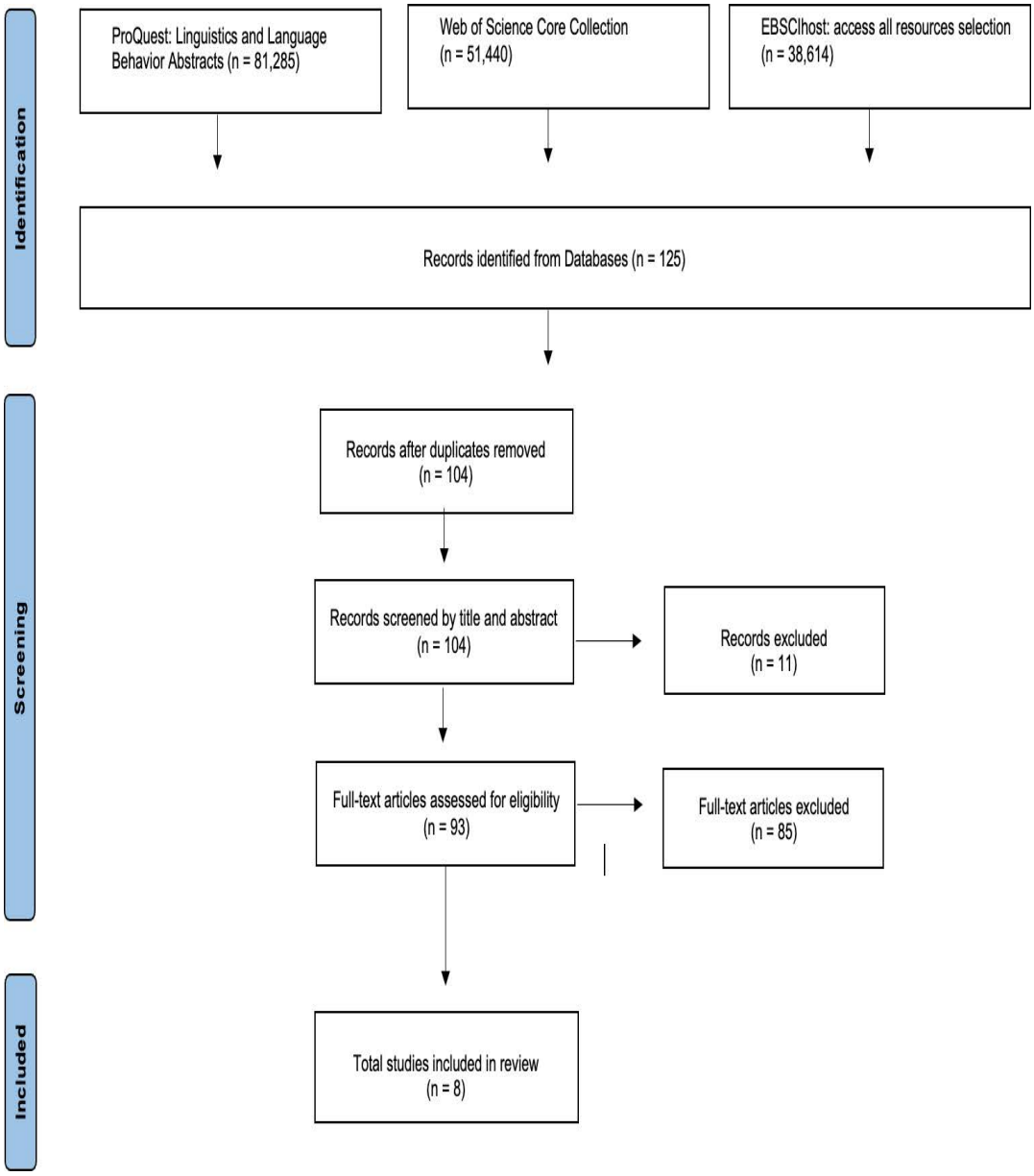


Figure 7.1. PRISMA flow diagram of search and selection process to identify suitable articles for inclusion in the scoping review

Following the database searches, 125 studies identified as potentially relevant were uploaded to Covidence (Covidence Systematic Review Software, 2024). Covidence is a web-based collaboration software platform that streamlines the production of systematic and other literature reviews. Of the 125 studies uploaded

to Covidence, 21 duplicates were identified and removed. Next, a final set of 104 studies were screened at the title/abstract level. Two reviewers independently screened titles and abstracts based on the pre-defined inclusion and exclusion criteria and 11 studies were deemed unsuitable and removed. No novel studies were located in the grey literature search. In the second search stage, the full text of the remaining 93 studies was evaluated against the inclusion and exclusion criteria by two independent reviewers. Conflicts at both the title/abstract and full text stages were resolved through discussion between reviewers. Eight articles met all inclusion and exclusion criteria and were ultimately selected. All of these studies were conducted in schools, kindergartens, and day-care centres, included children from different cultural and linguistic backgrounds, and examined the role of language, number naming systems, and numeracy experiences in children's early numeracy development.

7.9. Data Extraction

When articles were determined to meet eligibility criteria by the two reviewers, data were extracted by one reviewer, and the accuracy of extraction was independently cross-checked by the second reviewer. Data were extracted into an Excel spreadsheet (see Appendix), and organised according to the following categories:

- Author (Year)
- Citation
- Abstract
- Summary
- Methodology
- Purpose
- Research Questions
- Hypotheses
- Languages

- Participants
- Sample Size
- Assigned condition
- Mono-/Bilingual
- Setting
- Country
- Number Naming Variations
- Setup/ design
- Pre-/post-test
- Testing procedure
- Scoring
- Key findings
- Additional observations
- Conclusion
- Recommendations

A snapshot of the data extraction spreadsheet is provided below in Table 7.2. Given the size of the data extraction document, readers are referred to the Appendix for further information on the data extraction table.

Table 7.2. Snapshot of part of the data extraction table

etails	Study #1	Study #2	Study #3
uthor (Year)	Cankaya et. al (2014)	Clayton et. al (2020)	Dowker & Roberts (2015)
tation	Cankaya, O., LeFevre, J.-A., & Sowinski, C. (2014). The role of number naming systems and numeracy experiences in children's rote counting: Evidence from Turkish and Canadian children. Learning and Individual Differences, 32, 238–245.	Clayton, F., Copper, C., Steiner, A. F., Banfi, C., Finke, S., Landerl, K., & Göbel, S. (2020). Two-digit number writing and arithmetic in Year 1 children: Does number word inversion matter? Cognitive Development, 56, 100967.	Dowker, A., & Roberts, R. (2015). Does the transparency of the counting system affect children's numerical abilities? Frontiers in Psychology, 6, 945.
ostract	How important is the regularity of a number naming system when young children are learning to count? Three- and four-year-old English- and Turkish-speaking children (N = 75) played one of two versions of a board game. The number game involved counting and naming numbers from 1 to 20 whereas the color game involved naming colors. All of the Turkish-speaking children in the number game condition showed some improvement	Children's understanding of place-value structure is essential for later mathematics. Some languages (e.g., English) have an inversion property in two-digit names (e.g., "twenty-one" vs. "one-and-twenty"), which could influence children's number writing and arithmetic. In this study, Year 1 children speaking languages with and without inversion (English vs. German/Italian) were assessed on two-digit number writing and arithmetic	The transparency of number naming systems differs across languages, and this may influence children's numerical development. In this study, we compared children speaking languages with transparent counting systems (e.g., Turkish) and less transparent systems (e.g., English) on a range of numerical tasks, including counting, number comparison, and arithmetic. Results showed that transparency of the counting system was
ummary	The study conducted by Cankaya, LeFevre, and Dunbar (2014) investigates the influence of number naming systems and numeracy experiences on children's acquisition of numeracy skills, especially rote counting abilities, focusing on different levels of transparency of the Turkish and English number naming systems.	The study conducted by Clayton et al. (2020) examines the impact of number word inversion on the two-digit number writing and arithmetic abilities of 177 German-speaking and 309 English-speaking Year 1 children attending primary schools in Graz, Austria and Yorkshire, UK, respectively. The focus was on year 1 children, because they are in the early stages of learning two-digit numbers and arithmetic concepts. The mean age of the Austrian children was 7 years and	The study conducted by Dowker and Roberts (2015) explores the impact of the transparency of the counting system on children's numerical abilities. The researchers investigate whether the ease of mapping between number words and quantities in a counting system affects children's performance in numerical tasks. There are few studies available using a language other than Chinese as the transparent example, and the present study is the only one using Welsh. The

The data extraction table was used to organise the following aspects of data from the included articles: citations, summaries of articles, methodologies used, purposes of studies, research questions, hypotheses, languages of the studies. The following section describes the articles selected for inclusion in the scoping review.

7.10. Selected Articles

For ease of reference and reading, the eight studies identified in this scoping review will be henceforth referred by number as shown below in Table 7.3 which displays the selected articles study numbers, citations, year and titles. The general characteristics, similarities and differences in the dataset of eight articles, initially through a process described as charting the data are subsequently described.

#	Citation	Country / Language	Sample / Age	Study Type /Design	Notes
1	Cankaya, O., LeFevre, J.-A., & Sowinski, C. (2014). The role of number naming systems and numeracy experiences in children's rote counting: Evidence from Turkish and Canadian children. *Learning and Individual Differences, 32*, 238–245.	Turkey, Canada	N=75 children, 3–4 yrs	Experimental	Studied effect of number naming systems on rote counting
2	Clayton, F., Copper, C., Steiner, A. F., Banfi, C., Finke, S., Landerl, K., & Göbel, S. (2020). Two-digit number writing and arithmetic in Year 1 children: Does number word inversion matter? *Cognitive Development, 56,* 100967.	UK, Austria	N=486, 5–6 yrs	Cross-sectional, assessment	Investigated effect of number word inversion on writing and arithmetic
3	Dowker, A., & Roberts, R. (2015). Does the transparency of the counting system affect children's numerical abilities? *Frontiers in Psychology, 6*, 945.	Wales	N=40, 5–7 yrs	Experimental, Comparative	Explored transparency effects on counting, number comparison, Arithmetic
4	Helmreich, I., Zuber, J., Pixner, S., Kaufmann, L., Nuerk, H.-C., & Moeller, K. (2011). Language effects on children's nonverbal number line estimations. *Journal of Cross-Cultural Psychology, 42*(4), 598–613.	Austria, Italy	N=237, 5–8 Yrs	Experimental	Nonverbal number line estimation task; language effects
5	Lê, T. P., & Noël, M.-P. (2020). Transparent number-naming system gives only limited advantage for preschooler's numerical development: Comparisons of Vietnamese and French-speaking children. *PLOS ONE, 15*(12), e0243472.	Vietnam, Belgium	N=208, Preschoolers, 4–5 yrs	Comparative, assessment	Transparent number naming gave slight counting advantage
6	Li, X., Sun, Y., Baroody, A. J., & Purpura, D. J. (2013). The effect of language on Chinese and American 2- and 3-year olds' small number identification. *European Journal of Psychology of Education, 28*, 1525–1542.	China, USA	N=24, 2–3 Yrs	Experimental	Small number identification; English children performed better on "two," "three," "four"
7	Magargee, J., & Beauford, K. (2016). Do explicit number names accelerate pre-kindergarteners' numeracy and place value acquisition? *Educational Studies in Mathematics, 92*, 179–192.	USA	N-101, Pre-K children, 4–5 yrs	Longitudinal intervention	Explicit number names accelerated numeracy, especially with 2 yrs intervention
8	Villarroel, J. D., & Miñón, M. (2011). The origin of counting: A study of the early meaning of "one," "two," and "three" among Basque- and Spanish-speaking children. *Educational Studies in Mathematics, 76*, 345–361.	Basque Country, Spain	N=131, Preschoolers, 3–5 yrs	Experimental	Language & culture shaped understanding of "one"–"three"

Table 7.3. Articles selected for inclusion

7.11. Charting the Data

Charting is the process of extracting and synthesising key information from each included study in a structured, consistent way. The fourth step in Arksey and O'Malley's methodological framework, it allows researchers to map the characteristics of the evidence base, rather than just listing studies. In this stage of a scoping review, data are 'charted' from the studies included in the review. Charting involves synthesising and interpreting the data by sifting, categorising, and sorting material according to key issues and themes (Arksey & O'Malley, 2005, p. 22). A narrative review is provided below which extracts key contextual and empirical information from each study.

Thematically, of the eight publications, six dealt with the topic of transparent number names whilst two addressed the semantic singular/plural distinction. Two studies (#6 and #8) dealt with the possible benefits of semantic singular/plural distinction and six articles focused on the transparency of the language naming system itself. The two studies which examined the semantic singular/plural distinction aimed to comprehend cardinality, especially the ability to distinguish sets of "one" from sets of other small numbers. Both studies came to the same conclusions, that the semantic singular/plural distinction does indeed benefit children in comprehending cardinality.

The six articles (#1, #2, #3, #4, #5, #7) which addressed the putative benefits of transparent number names came to four different conclusions. This divergence is largely due to focus on different aspects of transparency and different testing methods.

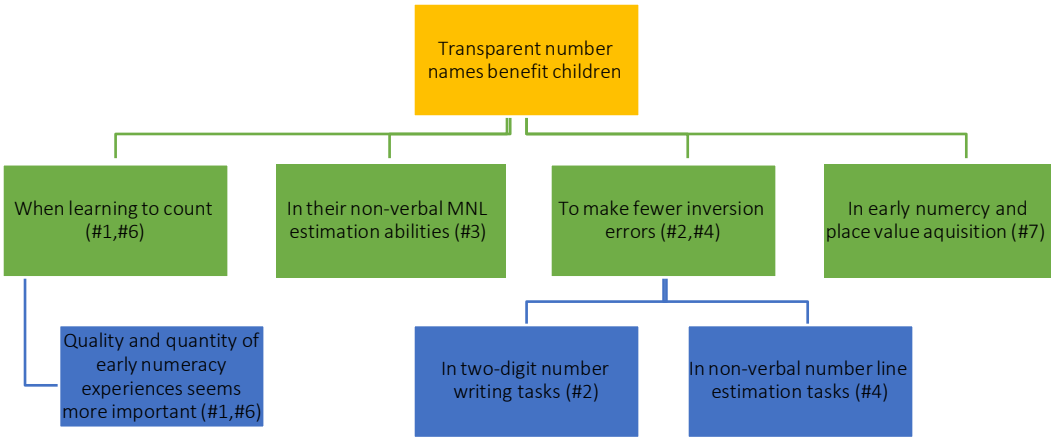


Figure 7.2. Findings regarding number name transparency

The eight studies discussed in this review were conducted in several different countries and featured different languages (Figure 7.3). Geographically, the studies are spread across the following locations:

- Austria
- Belgium
- China
- Italy
- North America (Canada, USA)
- Spain
- United Kingdom (England, Wales)
- Vietnam

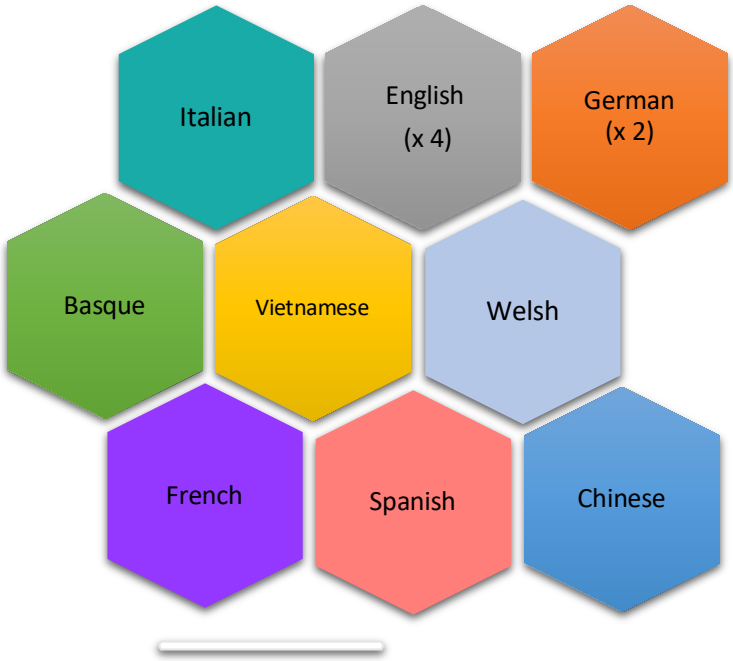


Figure 7.3. Languages included in studies

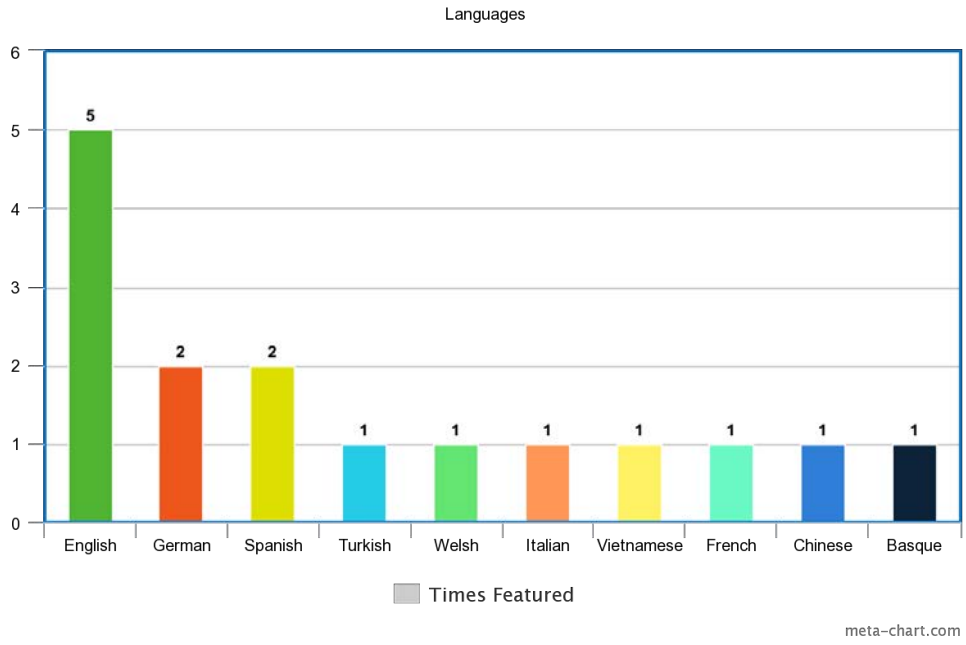


Figure 7.4. Occurrence of relevant languages

English is the language of focus in four studies, German appears in two studies, and French, Spanish, Basque, Vietnamese and Mandarin Chinese feature in single studies. A wide variety of languages are mentioned in the context of the studies (Figure 7.4.).

In terms of date of publication, no exclusion criteria were set. It is noteworthy that all located articles on this topic were published in 2011 or later, which reflects the start of a period of research activity.

Six of the studies were purely observational (#2, #3, #4, #5, #6, and #8) while two were intervention studies (#1 and #7). All participants in the eight studies were young children between the ages of 2 and 7 years old. The settings ranged from the pre-K setting to kindergartens (in five studies) and the primary school setting (twice). Sample sizes varied greatly, from 24 to 468 subjects, with five studies having a sample size greater than 100 subjects (Figure 7.5.).

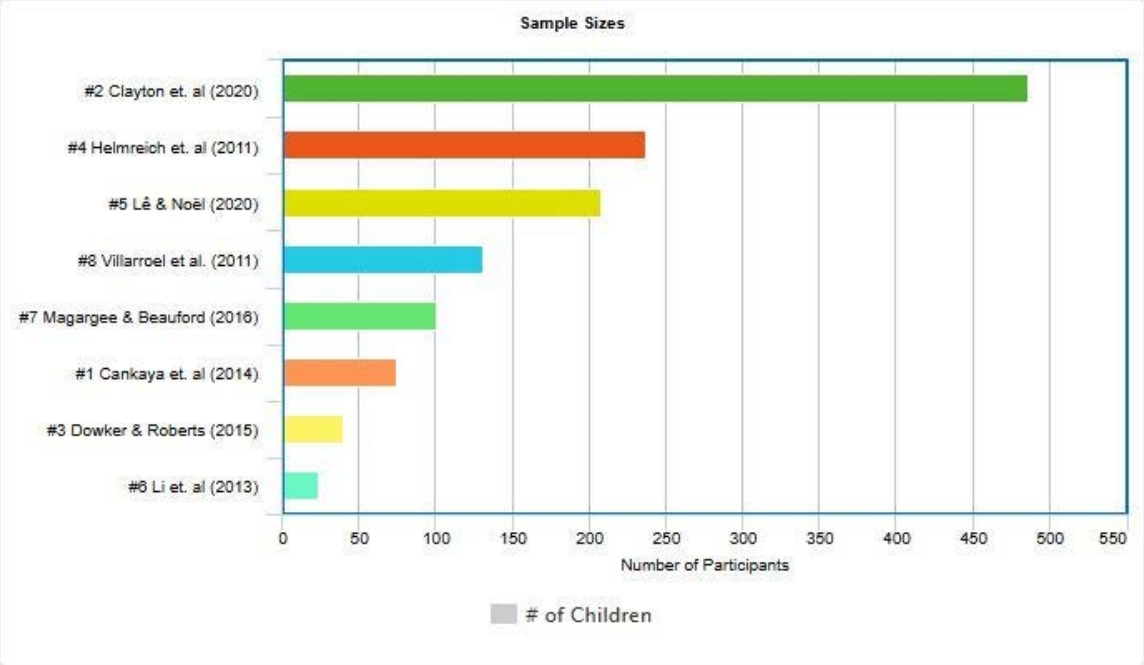


Figure 7.5. Difference in sample sizes across all featured studies

Regarding the focus of the research studies, the number naming variations investigated were explicit vs less explicit number naming in six of the studies, with two of these six focussing on invertedness, and the influence of (lack of)

semantic singular/plural markers on number learning in two of the studies, with a third mentioning it as an additional point in their findings. This is presented in Figure 7.6 below.

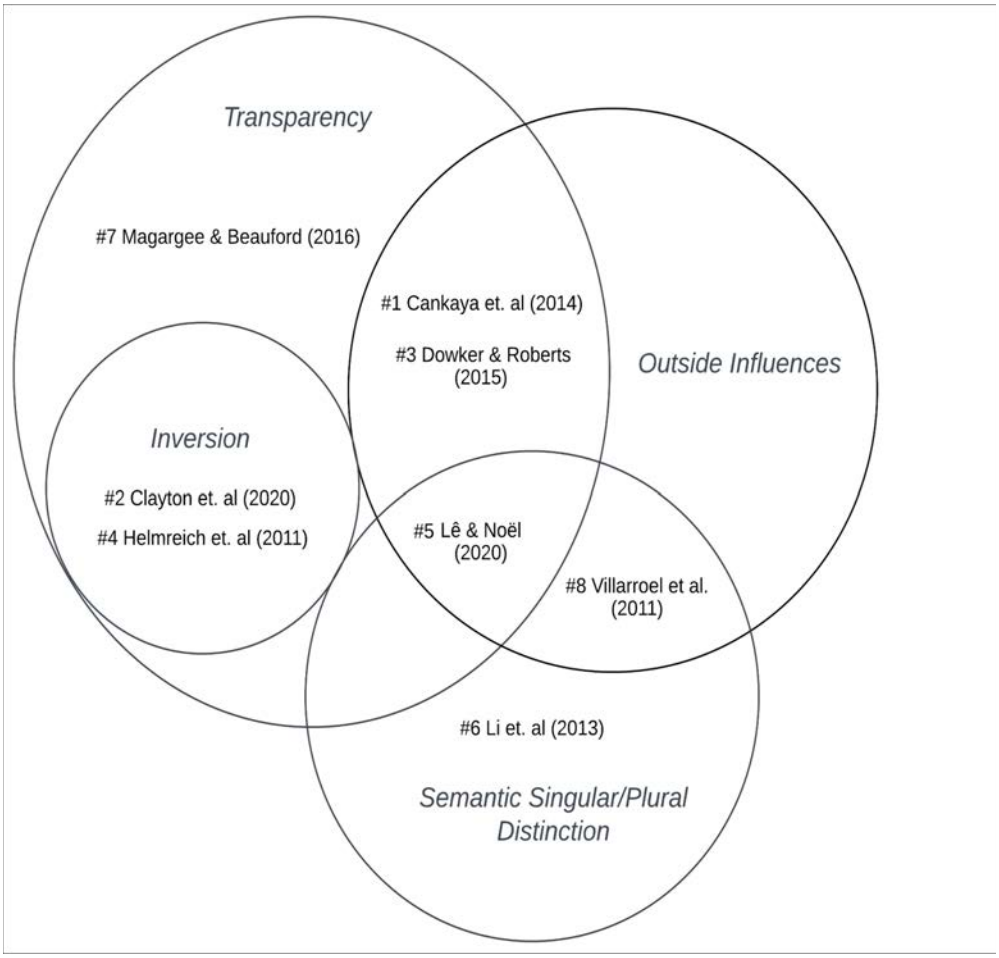


Figure 7.6. Areas of thematic overlap within the featured studies

The eight studies had three main areas of overlap, plus one subcategory. Of the six articles (#1, #2, #3, #4, #5, and #7) focussed on transparency:

- only #7 dealt solely with this topic in general
- #2 and #4 investigated inversion, a subcategory of transparency
- #1, in addition to transparency, touched upon the effect of outside influences on learning
- #3 made outside influences part of their experimental design by excluding differences in school system and cultural environment as best as possible

- #5 explicitly looked at transparency and outside influences in the form of home numeracy, and mentions that one of the languages used lacks semantic singular/plural distinction, which might account for one of their results
    - #6 and #8 focussed on semantic singular/plural distinction
    - #6 investigated only the effect of the lack of this distinction on children's early grasp of cardinality
    - #8 additionally indirectly addressed outside influences by conducting the study in a homogenous educational and cultural environment.

Most of the studies focussed on different aspects of number learning and the acquisition of numeracy in general:

- Influence of transparent number names on the acquisition of numeracy skills with emphasis on rote counting abilities (#1)
- Two-digit number writing and arithmetic abilities (#2)
- Non-verbal number line estimation (#3, #4)
- Transcoding (#3)
- Effect of transparent number names on the acquisition of various numerical skills (#5, #7).
- Effect of the lack of semantic singular/plural distinction on children's early grasp of cardinality (#6, #8)

All eight studies used a range of general tests including general cognitive and numerical tasks such as number writing, rote counting, and naming of digits. A diversity of testing methods of the studies were selected, including:

- Non-verbal number line estimation (#2, #3)
- Give-N and How-Many Tasks (#6, #8)
- Number and colour game (#1)
- Numerical tasks (#2, #5)
- Longitudinal study that used indirect testing through a general achievement test which was administered repeatedly over the span of six years (#7).

7.12. Findings

This section of the scoping review chapter conducts step 6 in Arksey and O'Malley's review stages, collating, summarising, and reporting the results. The sections below analyse and present the findings, firstly looking at the dataset of eight articles as a whole, and then providing a deeper dive into each of the eight studies for completeness.

7.13. Empirical Foundations

Turning first to commonalities in empirical foundations, of the eight studies included in the scoping review, four (#1, #3, #5, and #7) reference earlier research showing that transparent number-naming systems make learning to count easier for children.

Study #1 cited an experiment by Miller et al. (1995), which compared the early numeracy abilities of 99 pre-schoolers from the United States and China, aged three to five. Fifty percent of English-speaking children could not yet count to 20, while only twenty-five percent of Chinese-speaking children were unable to do so. However, there were no linguistic differences in the children's object counting and cardinality understanding, indicating that not all components of early numeracy are influenced by number naming. Due to the regularity of the Chinese number naming system, Miller et al. concluded that Chinese-speaking children require significantly less time to grasp the counting sequence than their English-speaking peers.

Study #5 also cited Miller et al. (1995), highlighting that Chinese children can

often count up to 40 at ages four to five, whereas American children of the same age reach this milestone about a year later. There were no differences in counting to 10, but fewer American children could count to 20. This framework argues that language traits related to the numerical lexicon, such as transparency, numerical morphology, or phonological and syntactic features, impact on numerical development. By contrast, inconsistent systems, like English or French, have number words that do not align perfectly with the Arabic symbolic system.

Study #7 cited neurocognitive research (Van Oers, 2001; Dehaene et al., 2004; Feigenson et al., 2004; Holloway et al., 2010), highlighting verbal and non-verbal processing of numbers.

Children's language centers store number names, which are linked to physical quantities and symbolic representations (Sousa, 2008). Fewer vocabulary words reduce memory load, facilitating the connection between number names, symbols, and quantities. Symbolic representations are read left to right in descending place value. In particular, Study #7 cited Sousa (2008), noting the brain's natural capacity to understand numbers as quantities rather than words. The Chinese number naming system explicitly expresses quantity and place value, with place values decreasing from left to right, mirroring the Arabic system. This facilitates faster learning of place value concepts in Chinese children compared to American children.

Study #3 referred to international comparisons of children's arithmetic performance (Fuson & Kwon, 1992; Miller et al., 1995; Miura & Okamoto, 2013), which have been attributed to the degree of correspondence between spoken and written number systems. Transparent counting systems may support

learning higher numbers and understanding place value concepts.

7.14. Novel Approaches

The studies also demonstrated novel approaches to the study of number names and numerical development. Study #1 noted that earlier research on the effects of number naming and children's number language left questions about the similarity or variance in children's numeracy experiences unanswered. The authors proposed controlled exposure to numerical skills (such as structured board games) to compare children's experiences. Study #3 suggested that differences in arithmetic or transcoding may arise from how counting systems are internally represented or how easily procedures are carried out in different systems. Few studies examine how transparency affects children's internal representations of numbers. The development of linear numerical magnitude representations, replacing logarithmic ones, occurs between ages 5 and 8, starting with smaller scales (Booth & Siegler, 2008).

Study #5 explicitly included home numeracy in its design. It concluded that cross-cultural differences between Asian and Western children could reflect schooling, mathematics exposure, or parent–child stimulation, in addition to transparency of the number naming system. East Asian children generally receive more practice and instruction, and mathematical ability correlates with parent–child numeracy activity frequency. Study #7 referred to the place value construct, using the three-level framework of Jones et al. (1996): Level 1 (pre-place value) uses single-digit numbers; Level 2 (composite unit understanding) represents initial place value knowledge; Level 3 (developing place value) applies

two-digit understanding for mental arithmetic. Children using explicit number naming systems typically enter at Level 2 and progress to Level 3 by the end of first grade.

Novel approaches were apparent in the study of inversion. Study #2 used verbal transcoding methods, employing semantic and asemantic models. Semantic models convert number words to magnitude representations before Arabic digits (McCloskey et al., 1985; McCloskey, 1992; Power & Dal Martello, 1990), showing numerical size effects. Asemantic models bypass magnitude activation (Verguts & Fias, 2006), with ADAPT (Barrouillet et al., 2004) representing spoken numbers phonologically and filling digit slots in memory.

7.15. Thematic Overview

The thematic overview below is organised by four central themes in the research foci: number naming variations, inversion, semantic singular/plural distinction, and outside influences. At this stage in the thesis, this section aims to provide an overview of the breadth of the literature located in the study, rather than a synthesis.

7.15.1. Number Naming Variations

The number naming variations investigated were transparent/explicit vs less transparent/explicit number naming in six of the studies, with two of these six focussing on invertedness, and the influence of (lack of) semantic singular/plural markers on number learning in two of the studies. The articles focussing on the effects of transparency of number naming systems were:

- #1 Turkish (transparent) vs. English (not transparent)
- #2 German (inverted) vs. English (not inverted)
- #3 Welsh (transparent) vs. English (not transparent)
- #4 Italian (not inverted) vs. German (inverted)
- #5 Vietnamese (transparent) vs. French (not transparent)
- #7 (artificial transparent NNS in English and Spanish).

Four of the studies (#1, #2, #4, #5) were conducted with the children of each language group living in different cultures and school systems while the subjects in two studies (#3, in Wales, UK, and #7, in Texas, USA) all lived within the same cultural context and attended schools within the same system.

Testing and interaction with the subjects were conducted individually in four of the studies (#1, #3, #4, #5) and in classroom settings (#2, #7). The focus of four of the studies (#1, #3, #5, #7) was on overall transparency of the number naming system as single influence on numerical abilities, while two publications (#2, #v4) focussed explicitly on inversion.

7.15.2. Inversion

The two articles focussing on inversion as a specific non-transparent number naming property were #2 (German (inverted) vs. English (not inverted)) and #4 (Italian (regular) vs. German (inverted). While study #4 investigated only whether these numbers had an effect on nonverbal number line estimation, study #2 administered four tasks, namely number writing, arithmetic, non-verbal reasoning, and working memory. Both studies were conducted among primary school children in Year 1 of schooling, and in both studies one group of children, the German-speaking inversion group, was on average significantly older that the non-inversion group, with better numerical abilities correlating with age in general, but poorer performance overall on the number line estimation task due

to making more inversion errors.

Study #2 highlighted that integrating tens and units correctly is challenging when number words do not transparently map onto Arabic place value. Research on number writing in primary school students (Dowker et al., 2008) shows that inversion complicates transcoding. Children using inverted number terms make more errors, making this a key focus for developmental models. Study #4 examined variations in place-value implementation across languages.

Language-specific characteristics, such as inversion, significantly affect German-speaking children's transcoding abilities (Zuber et al., 2009). Study #4 used non-verbal number-line estimation, relying on the mental number line (MNL) for number magnitude representation. Dehaene (2011) proposed a logarithmic-ruler-representation for numbers, while alternative theories suggest a shift from two-linear to linear coding for certain ranges (Ebersbach et al., 2008; Moeller et al., 2009).

7.15.3. Semantic Singular/Plural Distinction

Two of the articles (#6, #8) focused on the effect of the lack of semantic singular/plural distinction in a language on the acquisition of counting abilities and the understanding of cardinal meaning. Study #5 also mentioned the absence of singular/plural distinction in Vietnamese as a probable reason why fewer Vietnamese-speaking children were able to grasp the cardinal meaning of "two" as distinct from "one," although it did not focus on or elaborate further on this point.

Studies #6 and #8 reported that children from language backgrounds without a

singular/plural distinction struggled to differentiate the cardinal meaning of "one" from other numbers. Studies #5, #6, and #8 used the Give-N/Point-to-N task and found that the absence of singular/plural markers negatively affected children's acquisition of the cardinal meaning of the smallest numbers, particularly the distinction between "one" and "two."

Studies #2, #4, and #6 reached a similar conclusion, that the lack of singular/plural distinction delays a child's understanding of the cardinal meaning of "one" as distinct from other numbers. Study #6 further described an additional difficulty in Chinese due to the existence of different names for different meanings and contexts of "two." Specifically, Study #6 adapted the Point-to-N task using realistic objects, testing numbers 2–4 across multiple rounds, with correct responses defined as selecting the appropriate subset without counting. Study #8 employed the Give-N task, asking children to select n objects from an assortment and subsequently n+1 objects, classifying performance into no-number-knower, 1–3 knower, or CP-knower categories.

Both studies referenced Sarnecka et al. (2007), who found that Japanese children took longer than English or Russian children to distinguish "one" from "many," due to the absence of singular/plural marking in Japanese. Consequently, Study #6 concluded that singular/plural markers accelerate early number word learning, while Study #8 found that Spanish-speaking children outperformed Basque-speaking children in cardinality tasks, highlighting the influence of singular/plural distinctions on early numerical development.

7.15.4. Outside Influences

Five studies (#1, #3, #5, #7, #8) dealt explicitly or implicitly with the effect of

outside influences like culture, school system, or home numeracy. Studies #3, #7, and #8 were all conducted with participants within the same cultures and school systems (Spain, Wales (UK), and USA (Texas), respectively) and all found positive influence of number naming transparency on early numerical abilities independent of outside influences.

Studies #1 and #5 were each conducted in two countries and cultures (Vietnam/Canada; Turkey/France). Both studies came independently to the conclusion that home numeracy and early mathematical education in kindergarten have more impact on children's numerical ability than the regularity of the number naming system.

7.16. Consulting with Stakeholders

The final part of the scoping review method was to engage in a small number of stakeholder interviews, which are an optional sixth and final step in the Arksey and O'Malley (2005) framework. I elected to undertake this optional consultation exercise and attempt to consulting with stakeholders in order to refine the findings. I aimed to use the opportunity to meet with stakeholders, in this case educational practitioners, to share and explore together the findings of the scoping review, and to seek additional insights beyond the literature. To this end, a number of short informal individual interviews were conducted online with five mathematics practitioners, and each was interviewed individually. A convenience sample was initially used to make contact with staff, students and alumni associated with Marino Institute of Education, an associated college of Trinity College Dublin, which provides initial teacher education, as well as staff

connected with mathematics teacher training at the School of Education in Trinity College Dublin and at Queen's University Belfast, with snowball sampling used to find new contacts.

Early years practitioners with experience in teaching mathematics through different languages and/or who were training future teachers in how to prepare for a diverse linguistic environment were invited. Five practitioners responded positively, all of whom taught within a English/Irish language context or Irish immersion environment.

After obtaining Research Ethics Approval (please see Appendix) from the School of Linguistic, Speech and Communication Sciences at Trinity College Dublin, the process of making contact with and inviting potential interviewees was conducted via two gatekeepers at Marino Institute of Education and Trinity College Dublin. Mathematics teachers and teacher trainers were invited by email to participate in a short online interview. The invitation email included a Participant Information Leaflet and a short PowerPoint presentation with central findings from the reviewed articles (reproduced below). The nature of this study produced a highly specialised participant pool in spite of a multi-channel outreach effort. The interviews were designed as means of exploring the findings of the scoping review through the perspective of teachers and teacher trainers. Consulting these key stakeholders added the voices of practitioners and complemented the thematic analysis of the empirical data. The use of this last step in the scoping review protocol allowed for enhanced understanding of number naming, mathematical ability, and multilingualism in children. The following information was presented to the five practitioners in three slides:

(1) In the literature, two of the conventions of number names in some languages that can lead to delay and difficulty in learning early numerical skills are:

-Transparency
"[…]the consistency of conformity between the spoken and the written number system (usually the Arabic number system) and the degree to which it gives a clear and consistent representation of the base system (usually base-10) used in the language" (Dowker et al., 2008; 525).

-Inversion
Inversion refers to the non-standard order of parts of number names which generates names like 'four-teen' instead of 'te(e)n-four' in line with the order of the numerals 14.
In your experience, among the children you have worked with, have you noticed these features in the children in your classroom?

(2) One semantic convention in some languages that can lead to delay or difficulty in learning early numerical skills is:

-Semantic singular/plural distinction
These languages lack an equivalent to the function the plural suffix '-s' in English fulfils
The absence of this property in a language can delay the acquisition of cardinality, that is the understanding that the last number stated denotes the complete number of objects in the whole set
Have you experienced this distinction in your classroom?

(3) Tentative findings

A more transparent number naming system seems to benefit children when learning to count and in early numeracy and place value acquisition.
Children whose language has inverted number names are more likely to make inversion errors (writing e.g., 57 instead of 75) when writing Arabic digits to dictation and during non-verbal number line estimation
.
The lack of the Singular/Plural Distinction feature in a language appears to lead to initial difficulty distinguishing between amounts of 'one' and 'two' or, more generally, 'one' and 'many'.

What do you think, do you agree or disagree?

No personal or identifying data were collected or stored during the interview processes, which were conducted individually, one at a time, via Zoom (video call). All names and other potentially identifying information were removed. The interviews were not recorded, and the content was recorded through verbatim notes taken by the researcher. The five interviews took place in late 2023/early

2024. Two interviewees were female, and three were male. It must be noted that, due to the small number of participants and their varied backgrounds, the views presented here are not necessarily representative of the wider population.

Table 7.4 Stakeholder Interview Participants

	Gender	Role	First language (L!)	Date	Length of interview
1	Female	Lecturer in Maths	Irish	Nov 2023	71 mins
2	Female	Student teacher in mathematics	Ukrainian	Dec 2023	28 mins
3	Male	Student teacher in mathematics	Polish	Dec 2023	17 mins
4	Male	Mathematics teacher, teaching through Irish in a Gaelcholáiste	Irish	Jan 2024	42 mins
5	Male	Mathematics teacher, teaching through Irish in a Gaelcholáiste	Irish	Jan 2024	52 mins

It became apparent during each interview that none of the interviewees were aware of any of the effects of number naming conventions as presented in my slides. None of the interviews had never noticed any of said effects in their own teaching practice. The findings were entirely new to all of the participants. However, all of the interviewees when on to talk about the role of outside influences on early mathematical learning, including teaching and learning in different linguistic environments. These outcomes are presented in the next chapter. The outcomes from the stakeholder interviews point to how little is known about the influence of number names even amongst mathematics practitioners.

The stakeholder interviews conducted as part of this review provided additional insights into these findings. As we have seen earlier in the thesis, counting and numbering in Irish varies according on whether counting people, objects, or on the fingers. Double-digit numbers in the basic cardinal number names are

inverted. Irish, like Basque and certain Asian languages, lacks singular/plural distinction when counting objects, i.e. that there is no plural marker added to the end of the noun, regardless of how many objects are discussed. Given that Irish has three counting systems, it is considerably less transparent than the numbering system used in the English language. The stakeholder interviews with mathematics practitioners were particularly insightful regarding outside influences. When discussing outside influences on children acquiring early numerical skills through the medium of the Irish language, the most noticeable point raised by the teachers and teacher trainers was the need to bear in mind that Irish-speaking children are not only all bilingual in Irish and English, but that many are multilingual.

According to one interviewee, in her experience it is not uncommon for immigrant families to send their children to Irish medium schools, and that overall, Irish is not spoken at home in the vast majority of cases, in Irish and in immigrant families alike. This in turn means that most children will, solely or additionally, have learned basic numerical concepts through another language by the time they start Irish-medium creche or elementary school. Basic mathematical terms are also usually known to Irish-medium pupils, but more specialised ones are taught only in Irish. Furthermore, as another interviewee put it, Irish is not considered “cool” amongst pupils and is therefore not used to communicate even at school outside the classroom. Therefore, the influence of English on this particular language group must be borne in mind.

On the other hand, teachers who want to teach in English-medium elementary schools in Ireland are required to be fluent in Irish, and the formal and informal use of Irish by teachers is encouraged during lessons. The most widely used

counting system is the one for objects and during the first year numbers are not taught in their written form, because, as mentioned above, spelling depends on context. Due to Irish language revitalisation efforts, there has been a substantial increase in enrolments in Irish-medium schools in recent years. New efforts have been made by one interviewee and her colleagues to develop Irish language maths teaching courses for teachers, however they found that it proves difficult to merely translating existing maths teaching material from English into Irish, as it can distort the meaning and cause unintentional linguistic ambiguities. All in all, however, none of the sources reported any noticeable delays or difficulties in the acquisition of numerical skills in children who received their earliest numerical education in Irish, compared to English-medium pupils. As the school system in which both groups of children are educated is the same, except for the language of instruction, and given the fact of how substantially more opaque the Irish number system is, this is noteworthy, as it could be argued that the outside factor of educational and societal influences offsets any possible linguistic drawbacks in the area of numerical acquisition.

7.17. Discussion of Findings

This final section returns to the research question identified for this scoping review: What does the literature indicate about relationship between number-naming conventions in different languages and number acquisition in young children?

In the methodological framework for this study, Arksey and O'Malley (2005) highlight that the emphasis should not be placed on the weight or quality of evidence, but rather on a narrative account of existing literature. Having

conducted a thematic review and individual analysis of each study above, it is clear that the results of this scoping review indicate that the literature indicates is a relationship between number-naming conventions in different languages and number acquisition in young children. This scoping review found evidence for a relationship between three number-naming conventions in different languages and number acquisition in young children: (i) transparency, (ii) inversion, and (iii) semantic properties related to singular/plural markers.

The first two main findings in this study are that separate aspects of number names in some languages can lead to delay and difficulty in learning early numerical skills, namely transparency and inversion. Transparency is the degree to which the language's number naming system provides a clear and consistent representation of the base system (usually base-10) as well as the consistency of conformity between the spoken and written number system (usually the Arabic number system). Inversion refers to the non-standard order of parts of number names which generates names like "four-teen" instead of "te(e)n-four" in line with the order of the numerals 14.

Turning firstly to transparency, the research identified in this scoping review suggests that a more transparent number naming system benefits children when learning to count and in early numeracy and place value acquisition. Secondly, in terms of inversion, children whose language has inverted number names are more likely to make inversion errors (writing, e.g., 57 instead of 75) when writing Arabic digits to dictation and during non-verbal number line estimation.

The third main finding of this scoping review is that semantic properties related to singular/plural markers affect numerical skills in young children, especially

cardinality. Linguistic differences related to the semantic singular/plural distinction can lead to delay or difficulty in learning early numerical skills. This finding suggests that this is the case in languages that lack an equivalent to the function the plural suffix '-s' in English fulfils. The absence of this property in a language can delay the acquisition of cardinality, that is the understanding that the last number stated denotes the complete number of objects in the whole set. The lack of this feature in a language appears to lead to initial difficulty distinguishing between amounts of "one" and "two" or, more generally, "one" and "many".

During the initial search and selection process to source studies for inclusion in this scoping review it became quickly apparent that this is a field of study that is still in its infancy. As such, the methods and experimental designs utilised in the studies were varied and approached the central unifying question(s) from different angles. Despite this, all selected studies found a link between the differences in number naming conventions and the pace and ease in which numbers and numerical concepts are acquired by young children. In the case of the studies focussing on the transparency of number naming, it emerged that transparent number names appear to benefit children in accelerating learning to count, early numeracy and place value acquisition, resulting in children making fewer inversion errors, and their estimation abilities on a number line are more accurate at an earlier age. In the case of semantic singular/plural distinction, children having this distinction in their languages are faster to comprehend cardinality, especially the ability to grasp the difference between sets of "one" and sets of the following small numbers. However, in the studies that took this factor into consideration, there also seemed to be a theme suggesting that outside

influences like home numeracy and societal attitudes to early learning are factors on par with, or perhaps even more important to, early numerical acquisition than linguistic properties.

In addition to the above-mentioned necessity for more in-depth research into the overarching topic of interrelatedness between linguistic and cultural and/or educational influences, there are also more specific questions worth addressing. This is not surprising as this field is new and there has been only a limited amount of research conducted on the topic to date. Not much attention has been paid to whether the participants were mono or bi/multilingual and even less attention has been paid to the question of how this might affect their number acquisition linguistically or if this would alter their learning context in other ways. An interesting question to ask would be, for example, if and how an inverted and a non-inverted languages would influence each other in the mind of German and Italian bilingual children for example. If these cohorts were asked to complete a MNL task, would their inverted German hinder or the non-inverted Italian aid their MNL estimating skills, or would their performances lie somewhere in between German and Italian monolinguals. Another interesting question would be how would a Danish-German bilingual child perform? Both German and Danish feature an extended inversion subsystem, but Danish has several additional in-transparencies. Danish number names are very complex and largely incongruent with the place value system of Arabic numerals, it has opaque number names where the names for a primitive element, the same element in the Teen-class, and the same element in the Ty-class do not correspond to each other and are most difficult for children to reconcile with the base- 10 value system. It might be due to the fact that from 40 onwards

Danish number names follow a base-20 number naming convention.

Would the heightened complexity of Danish amplify the tendency to make inversion errors that the German monolinguals? Would they generally be slower in reaching number knowledge approximated by linear representation? As Ejersbo and Misfeldt (2011) put it, the logic of the base-10 system gets lost within the Danish number naming convention, so that many number names must be memorised. As a result, learning to count, and by extension to estimate numbers correctly on a number line, should take substantially longer for Danish children. Would Danish in addition to already in-transparent German confuse bilingual children even more, therefore rendering it harder to gain estimation skills, especially in regards to inversion errors? What would happen when a child grows up multilingual and a third language is added? Would possible effects, positive and negative, on estimation abilities cancel each other out because processing three languages is a big mental load to begin with?

It is important to gain a better understanding of all of these considerations, not only to achieve a better understanding of language related numerical learning and processing but also to potentially facilitate more inclusive and effective early maths education. Given the ever-increasing rate of migration and the resulting fact that classrooms are becoming more and more linguistically diverse, it is imperative to make lessons and education as accessible as possible to children from all linguistic backgrounds. The findings of this scoping review contribute to the current state of the art and set the groundwork for the further research that is needed to inform best practice for the facilitation of accessible numeracy instruction for children form linguistically diverse backgrounds in their first few years of formal education.

7.18. Conclusion

The findings of this scoping review demonstrate that there is a connection between the ease and speed at which young children pick up numbers and numerical concepts and the variations in number naming practices. Children who have languages with more transparent number names learn to count, acquire early numeracy, and understand place value more quickly. They also make fewer inversion errors and have more accurate estimation skills on a number line at a younger age. Children that have a semantic singular/plural distinction in their language are better able to understand cardinality, particularly the distinction between sets of "one" and sets of the following small numbers. The idea that external factors like home numeracy and societal attitudes towards early learning are just as significant as, or even more significant than, linguistic characteristics in relation to early numerical acquisition, was also common theme in the studies that took this element into account. The importance of external factors was the major finding from the stakeholder interviews. The properties of transparency, inversion, and singular/plural distinction have been shown to have a marked influence on early acquisition of numerical abilities and concepts. As research on this topic area is still quite new, it is as of yet impossible to say what potential long-term effects estimation inaccuracies, due to inversion or delayed understanding of cardinality or due to lack of singular/plural distinction could have.

Additionally, the interactions between early learning, number names, and outside influences are not yet fully understood. The fact that even mathematics practitioners are not aware of these influences, and are therefore not well placed

to mitigate them, is both understandable and concerning. From anecdotal evidence, many native speakers appear to be generally unaware of the quirks in their own native numbering systems, but as shown in this review, this does not mean that these linguistic peculiarities do not influence their numerical and mathematical learning experiences from the very start. The results of this scoping review may go some way toward increasing awareness of the connection between the ease and speed at which young children pick up numbers and numerical concepts and the variations in number naming practices. This study and future research in this area have important real-world application beyond understanding the intricacies of number names in different languages.

This chapter has reviewed the eight research publications selected for inclusion in this scoping review. The findings demonstrate a connection between the ease and speed at which young children pick up numbers and numerical concepts and the variations in number naming practices. Children who have languages with more transparent number names like Welsh or Vietnamese, seem to learn to count, early numeracy, and place value more quickly. They also make fewer inversion errors and have more accurate estimation skills on a number line at a younger age. Children that have a semantic singular/plural distinction in their languages like English, German, or Spanish, are better able to understand cardinality, particularly the distinction between sets of "one" and sets of the following small numbers. The following chapter will discuss the methodology and findings of the systematic review

.

Chapter 8: Systematic Review (Phase 2)

8.1. Introduction

Following on from scoping review presented in the previous chapter, this chapter reports the findings of a systematic review undertaken to explore the identified evidence base in greater depth. The findings are synthesised to highlight and appraise literature relevant to number names and numerical development in the classroom. After presenting the methodology and describing the twelve studies included in the review, this chapter assesses the evidence for the beneficial effects of DI on mathematics achievement and student attitudes. The chapter's conclusion summarises these results to emphasise critical success factors including diagnostic accuracy and striking a balance between structure and autonomy, suggesting that DI successfully enhances results when given sufficient assistance.

8.2. Review Protocol

A systematic review protocol was developed prior to study screening to ensure transparency and reduce bias (Boland, Cherry, & Dickson, 2017). The protocol follows established standards for systematic reviews, including the PRISMA-P 2015 guidelines for reporting systematic review protocols and the PICO framework. PICO refers to Population, Intervention, Comparator, and Outcome (Richardson et al. 1995). The protocol provides a robust framework in order to conduct an evidence-based review. The following five steps were adopted:

1. Identify the research question, based on PICO

framework (Population, Intervention, Comparator, and Outcome)
2. Define eligibility criteria, specifying inclusion/exclusion criteria, study design, participants, interventions, outcomes, language, and publication date, ensuring consistency and transparency in study selection.
3. Search for relevant studies, through a systematic search strategy using multiple databases, keywords, and controlled vocabulary; document search terms, dates, and databases to ensure reproducibility.
4. Select studies and extract data, screening titles and abstracts, then full texts, against inclusion/exclusion criteria. Use a data extraction tool to chart key information such as author, year, methods, population, intervention, outcomes, and key findings.
5. Synthesise and report findings, conducting a narrative synthesis, highlighting patterns, gaps, and implications for research and practice, and considering any bias.

8.3. Research Question

The research question posed in the systematic review was:

> *What does the literature say about the effect of differentiated instruction (DI) compared to non-DI on mathematics achievement levels and attitudes toward mathematics in primary and middle school students?*

Turning firstly to Population (P): the systematic review targets studies among primary and middle school students, allowing for a focus on foundational numerical development. The Intervention (I) is based on research involving Differentiated instruction (DI), looking at studies which investigate various pedagogical approaches and strategies which adapt content and processes to students' individual needs. The Comparator (C) compares non-differentiated and traditional instructional approaches, allowing for an assessment of DI's specific impact. Finally, the Outcome (O) looks at mathematics achievement levels and

students' attitudes toward mathematics, both of which are measurable indicators of the effectiveness and engagement associated with instructional practices.

8.4. Search Strategy

The following databases were searched for peer-reviewed research published in English in order to locate pertinent sources for this systematic review. Trinity College Dublin library subscriptions were used to access these databases:

- Web of Science Core Collection
- EBSCOhost: the following databases were accessed:
  - Academic Search Complete
  - British Education Index
  - ERIC
  - APA PsycInfo
- Google Scholar and Elicit[2] for supplementary 'grey literature' search, to locate theses, reports, and other non-peer-reviewed sources that may not be indexed in traditional academic databases.

8.5. Search Strings

To ensure a thorough retrieval of relevant material, the search strategy for this systematic review was developed utilising a combination of keywords (free-text terms) and controlled vocabulary (standardised descriptions). To reflect differences in terminology across databases, these phrases were methodically concatenated using truncation symbols (*) and Boolean operators (AND/OR).

[2] Elicit is an AI-supported research tool developed by Ought which reduces the manual work of screening/extracting data from large numbers of papers and provides identification of non-indexed sources. It is not a replacement for database searching but is a useful complementary tool for a comprehensive search strategy.

S1 (Language Diversity / Multilingualism):

(language* OR multilingual* OR bilingual* OR “linguistic difference*” OR "dual language" OR "dual-language" OR "language diversity" OR "language variation" OR "language background" OR "language proficiency" OR "language minority")

S2 (English Language Learners):

("ESL" OR "ESOL" OR "ELL" OR "English language*" OR "English as a Second Language" OR "English to Speakers of Other Languages" OR "English language learner*" OR "English learner*")

S3 (Early Childhood Education):

("early learn*" OR "early childhood" OR ECEC OR "early childhood education and care" OR kindergarten OR kindergarden OR "pre-kindergarten" OR "prekindergarten" OR "pre-K" OR "preschool*" OR "early years" OR "childcare" OR "nursery school*")

S4 (Mathematical Learning):

(math* OR number* OR count* OR "mathematical learning" OR "number learning" OR "number naming" OR "numeracy" OR "numerical skills" OR "number concepts" OR "number sense" OR "mathematics education" OR "math skills" OR calculation*)

S5 (Diff Instruction):

(“differentiated instruction” OR scaffolding OR "pedagogical strategies" OR "adapt*)

S6 (Transparency):

(“number transparency” OR transparent OR inversion OR”base-10” OR irregular* OR “number nam*” OR “opaque language*” OR “decimal system”)

The review used restricted vocabulary (descriptors) in addition to free-text keywords to guarantee a consistent and accurate search across many databases. These descriptors are pre-defined phrases supplied by database indexers (such as the ERIC or APA PsycInfo thesauri) to categorise the major topic matter of a study. The search method was able to find pertinent literature that might have used different natural language to explain the same topics by focusing on these terms within specified Subject or Descriptor areas. This approach was particularly helpful for bridging gaps where initial keyword strings for Language Diversity or Transparency failed to provide adequate results, ensuring that the final 12 studies represent the most precisely indexed data within the 2016–2026 timeframe. The descriptors used were:

- Mathematics
- Mathematic instruction
- Differentiated Instruction
- Number concepts
- Oral language
- Written language

- Elementary school mathematics
- Grouping practices
- Language minority students
- Mathematical achievement
- Classroom observation

The same search strings were used to identify grey literature in Google Scholar and Elicit. The search strings and descriptors were systematically applied across the databases in various combinations to identify relevant studies. Initial combinations incorporating search strings S1, S2, and S6 did not yield relevant results and were therefore excluded from subsequent searches.

8.6. Inclusion and Exclusion Criteria

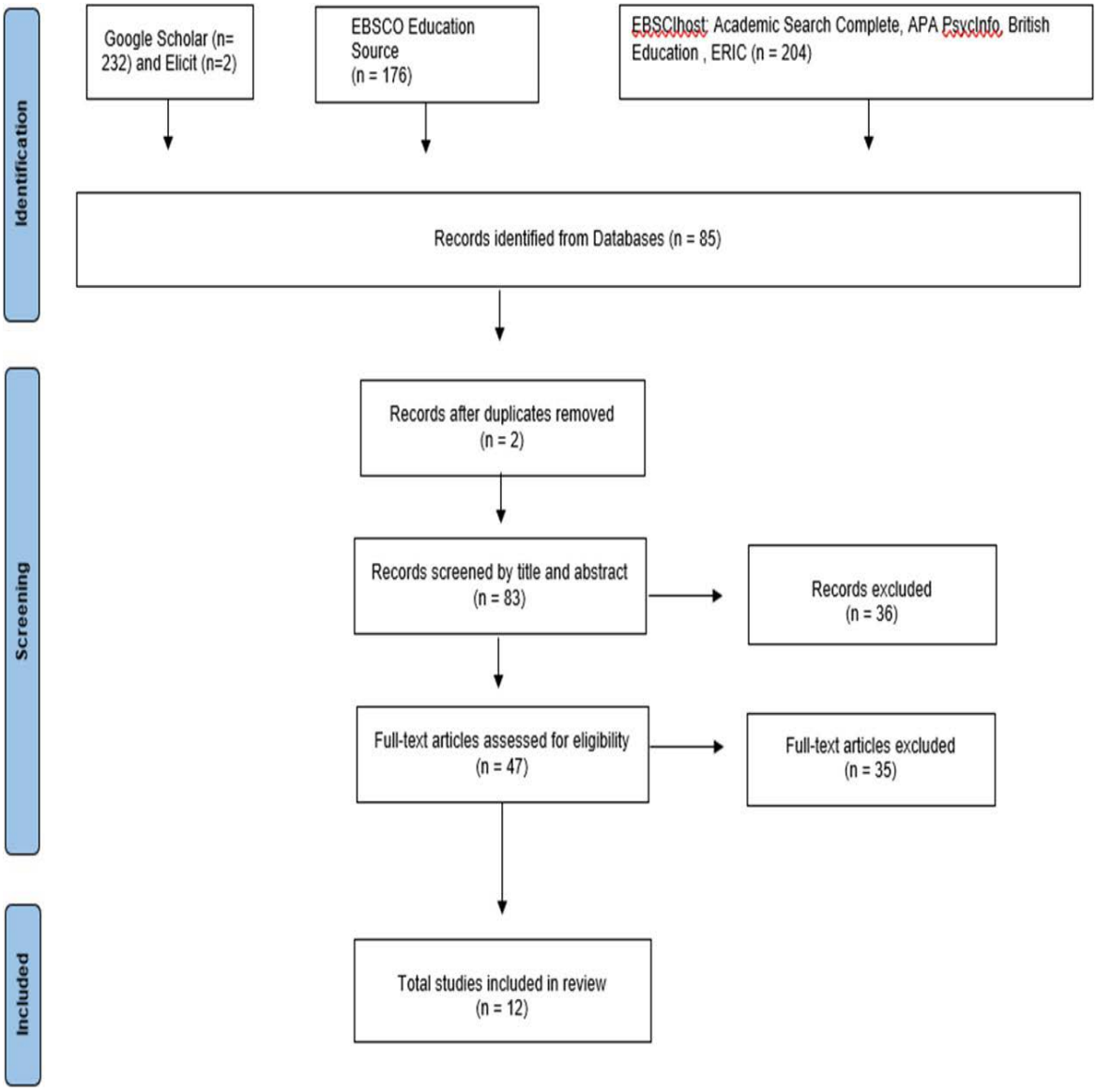
Unlike scoping reviews, systematic reviews set eligibility criteria in advance which clearly indicate what would and would not meet the review's inclusion criteria. The study adopted similar criteria to those developed during the course of the scoping review, although notably a time period was imposed, examining only literature published in the last ten years. The reason for this was that the scoping review indicated very little relevant research prior to this time period. The study's inclusion and exclusion criteria were as follows:

Table 8.1. Inclusion and exclusion criteria

Inclusion Criteria	Exclusion Criteria

• Original research published in:
  ○ peer reviewed books
  ○ conference papers & proceedings
  ○ dissertations & theses
  ○ scholarly journals
• Published in English
• Focus on primary and middle school students (ages 5-14 years) in general education settings
• Focus on implementation of differentiated instruction strategies in mathematics with clear description of DI approaches
• Studies that looked at the relationship between language and number system learning
• Published in in the last 10 years, between 2016 and 2026
	• Research conducted among child participants with autism, neurodiversity, and/or language impairment
	• Research conducted with deaf children and/or children learning and using sign language
	• Research that included or focused on adult learners/participants,
	• Research that focused on the involvement of parents or home learning
	• Involved machine learning or artificial intelligence

Figure 8.1. PRISMA flow diagram displaying search and article selection process



8.7. Study Selection

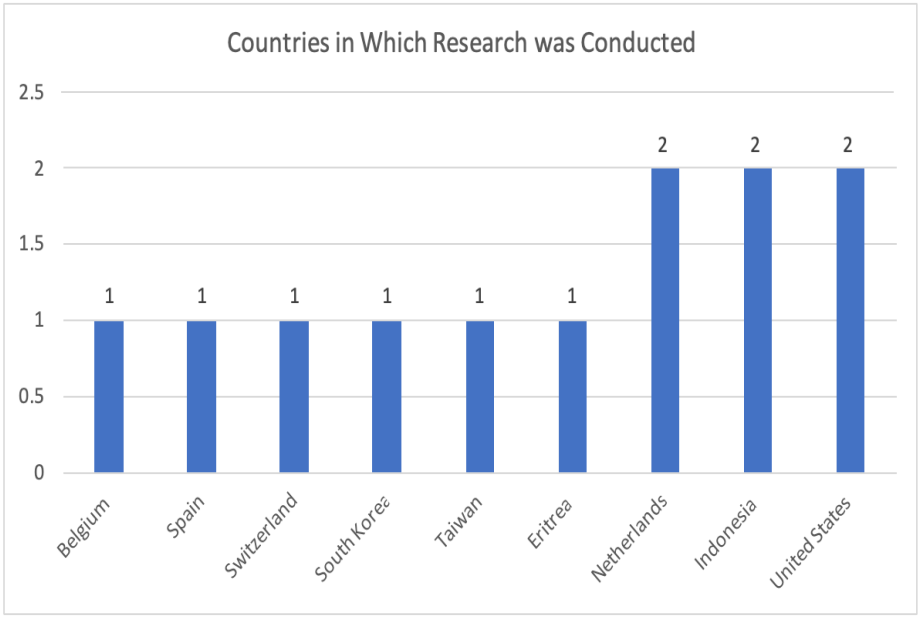
A total of twelve studies were selected for inclusion in this systematic review, displayed below in alphabetical order (Table 8.2).

Table 8.2. Studies selected for inclusion

Study No	Citation
Study #1	Ackermans, K., Bakker, M., van Loon, A. M., Kral, M., & Camp, G. (2025). Young learners' motivation, self-regulation and performance in personalized learning. *Computers & Education*, *226*, 105208.
Study #2	Akhsani, L., Junaedi, I., & Asih, T. S. N. (2025). Exploring insights in the creative thinking process of prospective mathematics teacher students. *Al-Jabar: Jurnal Pendidikan Matematika*, *16*(2), 581-596.
Study #3	Bang, H. J., Li, L., & Flynn, K. (2023). Efficacy of an adaptive game-based math learning app to support personalized learning and improve early elementary school students' learning. *Early Childhood Education Journal*, *51*(4), 717-732.
Study #4	Byeon, S., & Kim, N. (2020). Impact of Korean students' individual learning time on math performance: differential effect of teachers' assessment competency. *Asia Pacific Education Review*, *21*(4), 601-613.
Study #5	Febrilia, Y., Mardati, A., & Machbub, A. D. Z. (2025). Effects of differentiated instruction on primary students' mathematics achievement: A quasi-experimental study. *Al-Jabar: Jurnal Pendidikan Matematika*, *16*(2).
Study #6	Lai, C. P., Zhang, W., & Chang, Y. L. (2020). Differentiated instruction enhances sixth-grade students' mathematics self-efficacy, learning motives, and problem-solving skills. *Social Behavior and Personality: an international journal*, *48*(6), 1-13.
Study #7	Lavrijsen, J., Dockx, J., Struyf, E., & Verschueren, K. (2022). Class composition, student achievement, and the role of the learning environment. *Journal of Educational Psychology*, *114*(3), 498.
Study #8	Prast, E. J., Van de Weijer-Bergsma, E., Kroesbergen, E. H., & Van Luit, J. E. (2018). Differentiated instruction in primary mathematics: Effects of teacher professional development on student achievement. *Learning and Instruction*, *54*, 22-34.
Study #9	Rodríguez-Martínez, J. A., González-Calero, J. A., del Olmo-Muñoz, J., Arnau, D., & Tirado-Olivares, S. (2023). Building personalised homework from a learning analytics based formative assessment: Effect on fifth-grade students' understanding of fractions. *British Journal of Educational Technology*, *54*(1), 76-97.
Study #10	Zerai, D., Eskelä-Haapanen, S., Posti-Ahokas, H., & Vehkakoski, T. (2023). The use of question modification strategies to differentiate instruction in Eritrean mathematics and science classrooms. *Education Sciences*, *13*(3), 284.
Study #11	Zhang, D., Ding, Y., Lee, S., & Chen, J. (2017). Strategic development of multiplication problem solving: Patterns of students' strategy choices. *The Journal of Educational Research*, *110*(2), 159-170.
Study #12	Ziegler, E., Edelsbrunner, P. A., & Stern, E. (2021). The benefit of combining teacher-direction with contrasted presentation of algebra principles. *European Journal of Psychology of Education*, *36*(1), 187-218.

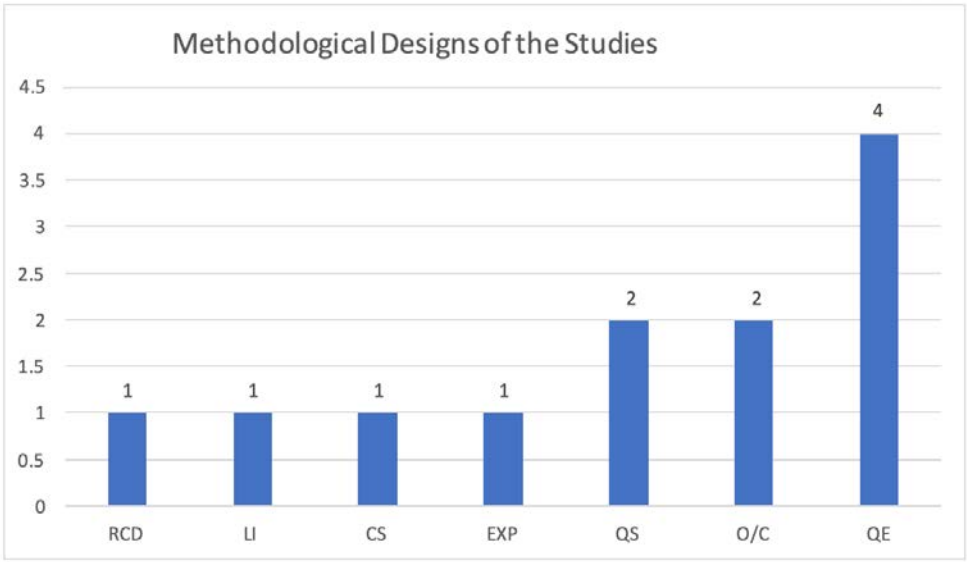
8.8. Characteristics of Included Studies

Figure 8.2. Countries in which research was conducted



The studies selected for inclusion were carried out in the Netherlands (#1, #8), Belgium (#7), Spain (#9), Switzerland (#12), Indonesia (#2, #5), South Korea (#4), Taiwan (#6), Eritrea (#10), and the United States (#3, #11), reflecting a significant geographic diversity.

Figure 8.3. Methodological Designs of the Studies



RCD: Randomized Controlled Design, LI: Longitudinal Pre-test–Post-test Intervention, CS: Cross-sectional Strategy Study, QS: Qualitative Studies, EXP: Experimental Classroom Studies, O/C: Observational/Correlational, QE: Quasi-Experimental

Researchers used randomised controlled designs (#3), qualitative descriptive investigations (#2). Quantitative methods included longitudinal correlational research looking at the effects of class composition (#7) and observational secondary data analysis using multilevel modelling (#4). Four researchers used quasi-experimental techniques, such as a large-scale longitudinal field studies (#1), a large-scale cohort design assessing teacher professional development (#8), a six-session controlled intervention (#9), and a non-equivalent control group design (#5). The methodology also includes cross-sectional cognitive strategy studies using Latent Class Analysis (LCA) (#11), experimental classroom studies intended to examine aptitude–treatment interactions (#12), qualitative classroom interaction studies using video recorded lessons (#10), and longitudinal pre-test – post-test intervention studies (#6).

The sample sizes ranged widely, ranging from 23–37 students in small-scale

studies (#2, #5) to almost 5,600 students in the Netherlands (#8) and nearly 3,000 students in Korea (#4) and Belgium (#7). Of the twelve studies, ten (#1, 3, 4, 5, 6, 7, 8, 9, 11, 12) presented quantitative results pertaining to mathematical achievement and five (#3, 6, 8, 9, 10) evaluated affective outcomes, including motivation, engagement, and self-efficacy.

The context in which DI was implemented had a substantial impact on the results of all the research. Large class sizes, such as 50–70 students per class in Eritrea, as well as a lack of resources and subpar equipment (study #10) and technological difficulties with student devices (study #9) were major obstacles. Study #1's longitudinal study was interrupted by the COVID-19 pandemic, which caused the focus to change to emergency remote teaching which may have had an impact on motivation and metacognitive outcomes. Another important issue that was identified in the findings was teacher preparation. According to study #8, ten three-hour team sessions in the Professional Development (PD) program and extra coaching training were necessary for teachers to deliver high quality DI implementation. Successful implementation of DI was found to depend heavily on teachers' knowledge of and proficiency with differentiation techniques (study #5). Teachers appreciated the adaptive platform's user-friendliness its capacity to adjust to the needs of students. However, treatment teachers occasionally utilised other interventions, contrary to instruction, which may have diluted the results according to study #3.

The number of resources needed varied greatly across studies, ranging from low-resource tactics like question modification that worked in large classrooms with limited resources (study #10) to technology-intensive approaches that required multidisciplinary design teams and ICT infrastructure (study #1).

Various professional development and training recommendations were made including differentiation strategies (study #2), focused professional development (studies #5, #8, #11), and taking local contextual factors into account when implementing DI approaches (study #1).

Table 8.3 Description of Methods of Selected Studies

Study Number	Study	Design	Sample	Data Collection	Analysis
Study #1	Ackermans et al. (2025)	Large-scale longitudinal quasi-experimental field study with between-subjects repeated measures; school-specific personalized learning interventions over 3 years	588 primary students (ages 7–12; Grades 2–6), 82 teachers, 8 Dutch elementary schools; students nested in classes and schools	National standardized achievement tests (math, spelling, reading, vocabulary); Junior Metacognitive Awareness Inventory; Intrinsic Motivation Inventory; student and teacher ICT-skills assessments	Bayesian multilevel Gaussian regression with random intercepts (students nested in classes/schools); MCMC estimation; Bayes factors and posterior distributions
Study #2	Akhsani et al. (2025)	Qualitative descriptive study of creative thinking processes in mathematical problem solving	Prospective mathematics teacher students (Indonesia; university level)	Numerical problem-solving tasks; written solutions; semi-structured interviews	Qualitative coding and thematic analysis aligned with stages of creative thinking (e.g., preparation, incubation, illumination, verification)
Study #3	Bang et al. (2023)	Quasi-experimental field study with treatment–control comparison in authentic classroom settings	986 early elementary students (Kindergarten & Grade 1): 505 treatments, 481 control; multiple schools in the U.S.	External standardized mathematics assessment: in-app adaptive assessment and log data (skills mastered, usage); teacher surveys	ANCOVA and regression analyses comparing learning gains; subgroup analyses by prior achievement; correlational analyses linking app usage to outcomes
Study #4	Byeon & Kim (2020)	Observational secondary data analysis using multilevel modelling	2,942 middle-school students, 426 math teachers, 207 Korean schools (GEPS panel data)	Panel survey data on individual learning time; math achievement scores; teacher feedback and assessment competency measures	Three-level Hierarchical Linear Modelling (HLM) (students nested in teachers and schools); cross-level interaction effects

Study #5	Febrilia et al. (2025)	Quasi-experimental non-equivalent control group design	Two intact Grade 4 primary classes (Indonesia): experimental vs. control	Validated mathematics achievement post-test	Assumption checks followed by independent-samples t-tests comparing post-test scores
Study #6	Lai et al., (2020)	Longitudinal pretest–post-test intervention study	25 sixth-grade students (Taiwan)	Mathematics self-efficacy scale; mathematics learning motivation scale; problem-solving performance Test	Paired-samples t-tests; mediation analysis using regression (self-efficacy → motivation → problem-solving)
Study #7	Lavrijsen et al. (2022)	Longitudinal correlational study examining class composition effects	2,895 Grade 8 students in 158 classes (Flanders, Belgium)	Standardized mathematics achievement tests; student background variables; teacher-reported instructional practices	Path analysis / Structural Equation Modelling (SEM) with mediation by learning-environment variables
Study #8	Prast et al. (2018)	Large-scale quasi-experimental cohort design evaluating teacher professional development	5,658 primary students (Grades 1–6) across 30 schools; three cohorts (two intervention, one control)	Standardized mathematics achievement tests administered longitudinally	Multilevel growth modelling (students nested in classes/schools); cohort comparisons
Study #9	Rodríguez-Martínez et al. (2023)	Quasi-experimental design with control group over six instructional sessions	127 fifth-grade students from Spanish public schools	Audience Response System (clicker) data; fraction-understanding tests; motivation questionnaires	ANCOVA and regression analyses comparing experimental and control groups
Study #10	Zerai et al. (2023)	Qualitative classroom interaction study	455 students and 8 teachers in 8 mathematics and science classrooms (Grades 4–6) in Eritrea	Video-recorded classroom lessons (11 lessons); lesson transcripts	Interaction analysis / qualitative content analysis identifying and categorizing question-modification strategies

Study #11	Zhang et al. (2017)	Cross-sectional cognitive strategy study	121 elementary students (Grades 3–5)	Written multiplication problem-solving tasks; coded strategy-use data	Latent Class Analysis (LCA) to identify strategy profiles; descriptive and comparative analyses of accuracy
Study #12	Ziegler, Edelsbrunner & Stern (2021)	Experimental classroom studies (two experiments) manipulating instructional approach	245 sixth-grade students (teacher-directed vs. self-directed conditions)	Pre-, post-, and delayed post-tests on algebraic principles	ANOVA and aptitude–treatment interaction analyses examining short- and medium-term effects

8.9. Findings

Having outlined the characteristics of the included studies, this section synthesises the main findings and organises them into five overarching themes: mathematical achievement, adaptive learning and personalisation, teacher-led differentiation, aptitude, and affective outcomes.

8.9.1. Mathematical Achievement

Of the twelve studies, eleven presented quantitative results pertaining to mathematical achievement. Teachers must constantly evaluate, categorize, and allocate tasks in order to implement manual differentiation. This can lead to an excessive cognitive burden. The findings of three studies, #1, #3, and #9, offer evidence that adaptive technologies can effectively close the divide between student preparedness and curriculum requirements, these studies indicate that technology addresses two of the most enduring challenges in differentiation, namely logistics and precision. While adaptive technologies, which guarantees that differentiation is regular and founded on detailed data, provides potent tools, studies #4, #5, #7, and #8 confirm that teacher-led differentiation is effective, while emphasizing the necessity of a high degree of professional expertise. The cognitive mechanisms of differentiation are examined in three studies, #2, #6, #11 and #12.

An analysis of the findings from these studies shows a consistent pattern. Differentiated Instruction reportedly had a consistently positive impact on mathematics achievement when compared to non-differentiated instruction. This overall conclusion, however, is qualified by considerations, such as, the delivery method (technology vs. teacher-led), the intervention's sustainability (study #8

showed notable impacts in Year 1 but not Year 2) and the students' specific cognitive profiles.

Table 8.4. Selected Studies that Reported Achievement Outcomes in Mathematics

Study Number	Article	Mathematics Domain	Direction of Effect
Study #1	Ackermans et al., 2025	• Measurements • Fractions • Ratios	Positive
Study #2	Akhsani et al. (2025)	• General mathematics (Mathematics achievement test)	Positive
Study #3	Bang et al., 2022	• Addition • Subtraction	Positive
Study #4	Byeon et al., 2020	• General mathematics (Mathematics achievement test)	Positive
Study #5	Febrilia we al., 2025	• General mathematics (Mathematics achievement test)	Positive
Study #6	Lai et al., 2020	• Algebra • Fractions	Positive
Study #7	Lavrijsen et al., 2021	• General mathematics (Mathematics achievement test)	Indirect positive effect
Study #8	Prast et al., 2018	• Numbers and number relations, covering the structure of the number line and relations between numbers • Addition and subtraction • Multiplication and division • Complex math applications, often involving multiple mathematical manipulations • Measuring (e.g., weight and length)	Positive Year 1, null Year 2
Study #9	Rodríguez-Martínez et al. 2023	• Fractions	Positive
Study #11	Zhang et al. 2017	• Latent Class Analysis (LCA)	Positive
Study #12	Ziegler et al., 2020	• Algebraic operations	Teacher-directed instruction leads to more positive effect

8.9.2. Adaptive Learning and Personalisation

Study #1 included a longitudinal analysis of personalized learning (PL) within Dutch primary schools. Their findings yielded very strong evidence that personalized learning interventions enhanced academic performance in mathematics. This research bolsters the scalability of technology-facilitated DI, indicating that when schools adopt personalization strategies in a systematic manner, the impact on achievement is considerable.

Study #3 examined how effective the adaptive, game-based learning application My Math Academy is for students in kindergarten and first grade in the U.S.A. Randomised Controlled Trial design was utilized in the study, ensuring that any differences observed could be linked to the intervention. The findings showed that the treatment group significantly outperformed the control group. The study uncovered a differential impact linked to developmental stage: the effect size for kindergarteners was larger than that for first graders. Importantly, the app was most effective on the most difficult skills that were evaluated, indicating that adaptive technology is especially useful for providing support with complex ideas that young learners may find challenging to understand in a conventional whole-class environment. This mechanism involves the algorithm's capability to deliver instant feedback and modify the challenge level in real-time, keeping the learner within their Zone of Proximal Development.

According to study #9, Learning analytics (LA) can be used to implement differentiation in older primary students (5th grade). Their investigation centred on students' comprehension of fractions, a domain that they stated is widely acknowledged to be challenging for students., The researchers made use of a

formative assessment grounded in learning analytics to create individualized homework tasks. Compared to their peers who received generic homework, students in the experimental group who were given tasks aimed at their specific error patterns, attained a significantly greater level of understanding. This discovery emphasizes an essential route for the effectiveness of DI: the close connection between diagnosis and intervention.

8.9.3. Teacher-Led Differentiation

Study #4 examined the connection between mathematics performance and Individual Learning Time (ILT) in a large-scale study of middle school pupils in South Korea. The authors discovered that although self-directed study time is generally advantageous, its efficacy varies. The study found that a significant moderator was teachers' assessment skill, notably the calibre of homework feedback. When combined with excellent, individualised instructor feedback, individual learning time produced noticeably greater achievement gains. On the other hand, the benefits of independent study decreased in the absence of this particular feedback. This reinforces the fundamental principle of differentiated instruction, which holds that ‘one-size-fits-all’ methods are ineffective and that students need focused feedback to make the most of their own study time. The results suggest that good differentiation goes beyond classroom instruction to include how teachers evaluate and react to students' homework.

Study #5 found that teacher-implemented DI significantly enhances achievement. Their investigation of Indonesian students in Grade 4 revealed that the class benefiting from differentiation tailored to readiness, interests, and learning profiles achieved a significantly higher score compared to the control

class. This discovery is crucial because it confirms DI in a context that might not depend on expensive technology, showing that the key to success is the pedagogical strategy of tailoring instruction to the child.

Study #7 investigated the connection between achievement and class heterogeneity in Belgium. According to their analysis, achievement was not adversely affected by class heterogeneity in and of itself. Nonetheless, the relationship between student outcomes and heterogeneity was mediated by individualised instruction. Higher achievement was associated with teachers using more differentiation tactics in more diverse classes. This finding is important from both a political and educational standpoint since it implies that high achievers do not suffer in mixed-ability classrooms as long as differentiation is done well. DI serves as a safeguard that permits diversity without sacrificing academic integrity.

Study #8 provides evidence for the positive impact of DI on mathematical achievement and also about the long-term viability of teacher-led differentiation. They investigated the impact of a teacher Professional Development (PD) program on DI in a large-scale study involving 5,658 students in the Netherlands. The study discovered a significant? positive impact on mathematics achievement for all ability levels (low, average, and high) in Year 1. In the second year, however, these effects diminished to nothing. This fade-out implies that maintaining high-quality differentiation is challenging when the intensity of intervention or training support diminishes. This suggests that DI is not a static skill that teachers simply acquire and retain, but rather a dynamic and demanding practice that necessitates continuous institutional support and coaching.

8.9.4. Aptitude: Treatment Interactions and Cognitive Profiles

The goal of study #2 was to explain and illustrate how maths students approach solving numerical issues. The findings showed that the steps of preparation, incubation, illumination, and verification are how creative thinking happens, and that individual cognitive styles and personality types strongly influence student performance. The researchers specifically discovered that guardian and artisan personality types were more likely to rely on previously taught methods and had difficulty coming up with novel ideas during the illumination phase, whereas rational personality types were able to comprehend difficulties more rapidly.

The psychological processes that connect academic performance and distinction are critically illuminated by study #6, This study, which looked at how DI affected the ability of sixth-grade students in Taiwan to solve mathematical problems, discovered that the experimental group fared noticeably better than the control group. The researchers discovered that a student's confidence in their own abilities, or Mathematics Self-Efficacy (MSE), was a significant predictor of success, explaining 77.6% of the variation in problem-solving scores. Differentiated instruction greatly raised students' self-efficacy, which in turn increased their learning motives (motivation), which in turn led to higher problem-solving performance, according to the study's mediation model. Students in the DI group reported feeling “more comfortable” and “assured” throughout tests, which was corroborated by qualitative feedback. This implies that DI enhances achievement by mending the learner's connection with mathematics rather than just changing the content.

Study #11 categorised American pupils according to how they solve multiplication problems using Latent Class Analysis (LCA). They distinguished two groups: those that relied on ineffective counting techniques or improper procedures, and those that used effective retrieval/algorithms. Six The implication for DI is that successful differentiation needs to focus on the student's strategy level. A completely different solution is needed for a learner who is stuck on unitary counting than for one who comprehends the idea but makes calculation mistakes. Study #12 investigated the relationship between algebraic teaching strategies and students' capacity for reasoning. They discovered a strong Aptitude-Treatment Interaction: whereas higher-ability students could handle less structured surroundings, pupils with weaker general reasoning skills profited considerably more from teacher-directed training combined with contrasting presentation of principles. The idea that DI is always synonymous with student-centred or inquiry-based learning is refuted by the researchers. According to the researchers, successful differentiation frequently entails giving lower-achieving children more structure and clear direction. In fact, these students may suffer if this structure is not provided in the sake of differentiation.

8.9.5. Affective Outcomes

Although most research in this review focussed on academic achievement, five studies (#3, #6, #8, #9, #10) specifically assessed affective outcomes such as self-efficacy, motivation, and engagement. The evidence indicates the beneficial effect of DI on students' psychological relationship with mathematics.

Table 8.5. Effects on Affective Outcomes

Study Number	Article	Measurement Method	Key Findings
Study #3	Bang et al., 2022	Teacher observation of engagement, motivation, confidence	Qualitative improvements in engagement and persistence observed
Study #6	Lai et al., 2020	Math self-efficacy (MSE), learning motives (MLM)	Significant improvement in MSE ($t(24)=-2.85$, $p<.01$) and MLM ($t(24)=-6.57$, $p<.001$)
Study #8	Prast et al., 2018	Teacher-perceived motivation	Teachers perceived positive effects on student motivation
Study #9	Rodríguez-Martínez et al., 2023	Reduced Instructional Materials Motivation Survey (RIMMS) (attention, relevance, confidence, satisfaction	Significant positive effects across all dimensions; $r=-0.43$ to -0.49
Study #10	Zerai et al., 2023	Qualitative video observation and interaction analysis of classroom discourse	Five distinct question modification techniques (repetition, rephrasing, clarification, decomposition, and code-switching) to be employed in large, resource-constrained classrooms.

Qualitative evidence of involvement in early childhood environments was offered by study #3. Teachers noted students counting, singing, and expressing excitement as they engaged with the adaptive math app. In the early years (Kindergarten/1st Grade), this form of engagement is essential, as it establishes a basis for a positive attitude toward school. Due to the app's adaptive nature, students seldom found themselves stuck or bored, thus preserving the flow state typical of high engagement.

Study #6 involved sixth-grade students in Taiwan and investigated the psychological mechanisms that support DI. Compared to the control group, students in DI classrooms reported significantly higher levels of Mathematics

Self-Efficacy (MSE) and stronger learning motives. They carried out a mediation analysis that showed learning motives to be a partial mediator of the connection between self-efficacy and problem-solving skills. Practices of differentiated instruction (for example, giving students tasks that are appropriately challenging) boost students' self-efficacy beliefs. This conviction drives the motivation to learn, which results in improved problem-solving performance. Students in the DI group explicitly stated that they felt more comfortable before tests and more assured when completing assignments. According to the researchers this may indicate that DI serves as a buffer against debilitating math anxiety that can often peaks in middle school.

In study #8, teachers were invited to complete teacher evaluation questionnaire regarding their Personal Development programme at the conclusion of Year 1 and Year 2, with 76 teachers responding at the end of Year 1 and 73 at the conclusion of Year 2. Teachers expressed moderate positivity regarding their learnings from the PD, as all responses exceeded the scale's midpoint. They also indicated that they had learned about all the stages in the differentiation cycle. In addition, most of the teachers in this study, 76.3% of those in Cohort 1 and 76.7% of those in Cohort 2, expressed partial or complete agreement regarding their actual use of what they had learned in the professional development for preparing and teaching mathematics lessons. Most importantly in the given context, teachers recognized beneficial impacts of the increased use of Differentiated Instruction on students' motivation and achievement.

Study #9 employed the Reduced Instructional Materials Motivation Survey (RIMMS) scale to assess motivation according to the ARCS model (Attention, Relevance, Confidence, Satisfaction). All motivational dimensions show notable

gains, according to the study's findings. A statistically significant movement toward more positive student views is indicated by the correlation coefficients, which varied from r = -0.43$ to r = -0.49. This statistical tendency suggests either a reduction in negative motivational emotions such as frustration or boredom or a matching increase in positive ones. This finding is particularly significant because as they pertain to the content domain of fractions which, according to the authors, have historically been linked to significant student frustration and a lack of engagement. They stated that the intervention likely removed the "frustration zone" (tasks that are too difficult) and the "boredom zone" (tasks that are too simple) by personalizing homework through learning analytics. Students' levels of confidence and satisfaction with their learning experiences grow naturally when they tackle tasks they view as difficult but doable. Study #10 offers insightful qualitative evidence of behavioural involvement in a setting with limited resources. The authors observed science and maths classes in Eritrea, where class sizes frequently exceed 60 students. They discovered that teachers employed question modification techniques, repetition, rephrasing, clarification, decomposition, and code-switching, to differentiate instruction in real-time. By lowering the social risk of providing an incorrect response in front of a sizable peer group, these language adjustments fulfilled an important affective purpose. Teachers were able to maintain a lively classroom environment, where students continued to actively participate in the shared conversation rather than retreating by lowering the threshold for involvement, through scaffolding and using the students' mother tongue. This study shows that differentiation was a motivating tool that fostered inclusion in a big, mixed-ability group. It created a supportive emotional environment in addition to being a useful cognitive tool for academic access.

It is important to note that seven out of the twelve studies included in this review did not measure attitudinal outcomes explicitly. This denotes a gap in the research. Should self-efficacy be viewed as a motivator for achievement, as proposed by study #6, then research that considers solely test scores fails to account for a crucial element in the explanation. Affective measures must be included in future research to determine whether achievement gains are associated with increased student stress or, as current evidence suggests, are a result of improved psychological well-being.

8.9. Discussion of Findings

The systematic review responded to the question: What does the literature say about the effect of differentiated instruction (DI) compared to non-DI on mathematics achievement levels and attitudes toward mathematics in primary and middle school students?

In order to address this research question above and to understand how and when differentiation works, it is important to look beyond the results of individual studies and consider the synthesised results. When considered collectively, the evidence suggests that, when compared to non-differentiated methods, differentiated instruction consistently produces beneficial benefits on students' attitudes and mathematical achievement.

Taken together, the studies identified above indicate that DI is a more effective pedagogical strategy than non-differentiated instruction for both academic achievement and student attitudes. DI achieves these outcomes by identifying specific cognitive demands, providing appropriate scaffolding, leveraging

technology to manage the cognitive load of individualised instruction, and enhancing self-efficacy. However, the fade-out effect observed in some teacher-led interventions suggests that a strong technological infrastructure or ongoing, integrated professional development for teachers is necessary for DI to produce sustainable, long-term benefits. Across the studies included in this review, positive effects of differentiated instruction on mathematics achievement were reported in the majority of cases amongst primary and middle school pupils. The positive outcomes seem to occur most strongly when DI is delivered through adaptive technologies (studies #1, #3), incorporated assessment-driven personalisation targeting particular skills (study #9) and is reinforced by ongoing teacher professional development (study #8). The benefits may be greater for younger pupils than for older students (study #3), and without continuous implementation support, the impacts may decrease (study #8). Notably, study 8 discovered notable impacts in Year 1 but not Year 2 of their 2-year study, which raises concerns regarding sustainability.

There is limited data on attitudes towards mathematics; only four studies having measured affective results. The outcomes that are available consistently show favourable effects on student engagement (study #3), motivation (study #9), and mathematical self-efficacy (study #6). According to study #6, learning motives partially mediated the association between self-efficacy and problem-solving abilities, indicating that improvements in attitude may be linked to advances in success. While more thorough research on attitudinal outcomes is required, the available research generally supports DI as effective for mathematics achievement in primary and middle school environments, with the highest results seen in well-supported, assessment-driven implementations. DI

implemented using adaptive technologies with personalised content consistently demonstrates small-to-moderate positive impacts on achievement in primary school mathematics (preschool–6th grade) (studies #1, #3, #9). At least in the short-term, professional development on differentiated instruction and within-class ability categorisation can have modest benefits (study #8). In classes with students who have varying levels of background, ability, and interest DI a positive relationship with achievement (study #7), and teacher feedback practices and formative assessment seem to moderate the relationship between individual learning time and achievement in middle school contexts (study #4). Although there are few controlled studies investigating affective outcomes, limited information on attitude outcomes consistently points to favourable impacts on motivation, self-efficacy, and engagement (studies #3, #6, #9). According to study #6, there is a positive correlation between problem-solving abilities and mathematics self-efficacy, which implies that improvements in attitude may partially explain gains in success.

The evidence from the systematic review show that the mechanisms underlying the effectiveness of DI vary greatly based on the form of delivery, learners' particular cognitive profiles, and the intervention's sustainability. DI enhances accomplishment by precisely tailoring the instructional format to learners' unique cognitive needs rather than just simplifying the information. The idea of differentiating only by ability level is challenged by evidence from study #11. Their Latent Class Analysis (LCA) showed that students fall into different qualitative profiles based on the use of strategies, such as retrieval versus counting, indicating that effective differentiation must target particular developmental obstacles rather than general competence.

Study #2 emphasised that students' creative thinking processes differ greatly, requiring differentiation that considers a variety of cognitive styles to support students to move from problem preparation to verification. This supports the need for cognitive alignment. For instance, study #12 showed a substantial aptitude-treatment interaction, defying the notion that difference usually entails open-ended (instructional tasks with multiple correct answers or various solution paths) or student-centred learning. In this study, the authors discovered that, while students with strong reasoning skills thrived in self-directed settings, organised teacher-directed instruction produced noticeably higher results for students with weak reasoning skills. This finding suggests that rather than merely slowing down explanations and instructions, effective differentiation for difficult students should take the form of more structure, clear guidance, and evidence-based strategies.

The difference between teacher-mediated and technology-mediated differentiation is one of the main themes that was identified from the cross-study analysis. Studies #1, #3 and #9 apply adaptive technology and claim strong positive benefits because technology seems to alleviate the logistical bottleneck of distinction by automating diagnosis and task assignment. For example, study #9 reported that fraction comprehension was greatly enhanced by utilising learning analytics to offer customised homework based on particular error patterns. Similarly, study #3 reported that adaptive algorithms worked especially well for kindergarteners. This is likely because the software offered the instantaneous, fine-grained scaffolding that is challenging for a single instructor to provide to a big class at once.

Teacher-led differentiation is clearly effective, especially for guaranteeing

access to learning for students of varying aptitude and motivation and so mediating class heterogeneity. According to study #5, teacher-planned differentiation, that is cognisant of student interest and readiness, works noticeably better than traditional instruction. Study #7 reported that DI serves as a important mediator in classes comprised of students with differing aptitudes and motivation, and that class variety does not negatively impact achievement as long as teachers actively distinguish between the different educational needs of students. While teacher professional development increased student achievement in the first year, study #8 noted that these effects diminished in the second year, underscoring the challenge of sustaining high-intensity differentiation without ongoing assistance. The findings demonstrate that DI is not limited to contexts with abundant resources. Study #10 showed how teachers in big, resource-constrained classes effectively varied instruction using question modification techniques including code-switching and decomposition. These strategies are useful for maintaining learners' interest in the group sessions, even though they place a greater emphasis on access and involvement than individualised worksheets.

Lastly, this analysis points to some psychological processes that connect DI to improved attitudes. According to study #6, DI is linked to increased mathematics self-efficacy, which enhances problem-solving abilities. DI lessens the tension that comes with challenging assignments by making sure they are within the student's Zone of Proximal Development (Vygotsky, 1978). Study #3 found qualitative increases in behavioural engagement among young learners using adaptive apps, and study #9 showed significant improvements in confidence and satisfaction.

8.10. Conclusion

The findings of the selected studies suggest that, as compared to non-differentiated education, DI has a positive, significant impact on attitudes toward mathematics and on mathematical achievement. This systematic review has demonstrated how differentiated instruction improves both mathematics achievement and student attitudes by bridging the gap between varied learner needs and curricular demands. The findings showed that, although technology provides effective tools for accurate diagnosis of students' educational needs, the effectiveness of differentiated instruction ultimately rests on sustainable implementation and careful alignment of instructional structure with students' unique cognitive profiles. The following chapter draws together the findings from both review phases, and considers their implications as well as future research directions.

Chapter 9: Discussion & Conclusion

9.1. Introduction

This final chapter brings the thesis to a close by synthesising the findings from the scoping review and the subsequent systematic review. It begins by restating the overall research aims and research questions, before providing an integrated summary of the key findings arising from both stages of the evidence synthesis. The chapter then outlines the contribution this thesis makes to knowledge in the field, followed by a discussion of the pedagogical implications of the findings and their implications for future research. The limitations of the thesis are subsequently acknowledged, and the chapter concludes with final reflections on the significance of the work and its contribution to understanding differentiated instruction in mathematics education.

9.2. Restatement of the Research Aims and Question

This thesis's principal objective was to examine the complex connection between early numerical development and language structures while also assessing Differentiated Instruction (DI) as a teaching strategy to address the resulting student variability. This study aimed to provide a thorough understanding of how language-specific number-naming conventions impact cognitive representations of numbers and how specific instructional strategies can mitigate potential disadvantages by bridging the gap between linguistic theory and educational practice.

A comprehensive assessment of the research done during the two stages of this

study is given in the next chapter. To provide context for the analysis that follows, it starts by restating the precise research topics for the scoping review (Phase 1) and the systematic review (Phase 2). An extensive synthesis of the results follows, combining the pedagogical solutions offered by DI with the linguistic barriers noted in early life. The chapter concludes by acknowledging the limitations of the current work, discussing the wider implications for policy and practice, and making suggestions for further research. The central research question that the scoping review set out to answer was:

> *What does the literature indicate about the relationship between number-naming conventions in different languages and number acquisition in young children?*

The results of the scoping review found a relationship between number-naming conventions in different languages and number acquisition in young children. The scoping review showed evidence for a relationship between three number-naming conventions in different languages and number acquisition in young children: (i) transparency, (ii) inversion, and (iii) semantic properties related to singular/plural markers.

The first two main findings in this study are that separate aspects of number names in some languages can lead to delay and difficulty in learning early numerical skills, namely transparency and inversion. Transparency is the degree to which the language's number naming system provides a clear and consistent representation of the base system (usually base-10) as well as the consistency of conformity between the spoken and written number system (usually the Arabic number system). Inversion refers to the non-standard order of parts of number

names which generates names like “four-teen” instead of “te(e)n-four” in line with the order of the numerals 14.

Regarding transparency, a more transparent number naming system benefits children when learning to count and in early numeracy and place value acquisition. Secondly, in terms of inversion, children whose language has inverted number names are more likely to make inversion errors (writing, e.g., 57 instead of 75) when writing Arabic digits to dictation and during non-verbal number line estimation. Semantic properties related to singular/plural markers affect numerical skills in young children, especially cardinality. Linguistic differences related to the semantic singular/plural distinction can lead to delay or difficulty in learning early numerical skills. This is the case in languages that lack an equivalent to the function the plural suffix ‘-s’ in English fulfils. The absence of this property in a language can delay the acquisition of cardinality.

After the scoping review's broader mapping of linguistic influences, a systematic review was carried out to offer a thorough, in-depth assessment of the efficacy of differentiated instruction. This review went beyond descriptive trends to synthesise high-level evidence on particular pedagogical outcomes. A comprehensive assessment of the research done during the two stages of this study is given in the next chapter. To provide context for the analysis that follows, it starts by restating the precise research topics for the scoping review (Phase 1) and the systematic review (Phase 2). The data are then thoroughly synthesised, the research and instructional implications are examined, and the limitations of the study are discussed.

The following research question was posed in the systematic review:

What does the literature say about the effect of differentiated instruction (DI) compared to non-DI on mathematics achievement levels and attitudes toward mathematics in primary and middle school students.

The evidence collected in the systematic review suggests that, when compared to non-differentiated methods, differentiated instruction consistently produces beneficial benefits on students' attitudes and mathematical achievement. However, beyond this key finding the synthesised evidence shows that the mechanisms underlying this effectiveness of DI vary greatly based on the form of delivery, learners' particular cognitive profiles, and the intervention's sustainability.

The findings indicate that DI enhances accomplishment by precisely tailoring the instructional format to learners' unique cognitive needs rather than just simplifying the information. The idea of differentiating only by ability level was challenged by evidence from one study (#11). Their Latent Class Analysis (LCA) showed that students fall into different qualitative profiles based on the use of strategies, such as retrieval versus counting, indicating that effective differentiation must target particular developmental obstacles rather than general competence. Also, students' creative thinking processes differ greatly, requiring differentiation that considers a variety of cognitive styles to support students to move from problem preparation to verification. This further supports the need for cognitive alignment. One study (#12) showed a substantial aptitude-treatment interaction, defying the notion that difference usually entails open-ended (instructional tasks with multiple correct answers or various solution paths) or student-centred learning. While students with strong reasoning skills thrived in self-directed settings, organised teacher-directed instruction produced

noticeably higher results for students with weak reasoning skills. This finding suggests that rather than merely slowing down explanations and instructions, effective differentiation for difficult students should take the form of more structure and clear guidance.

The difference between teacher-mediated and technology-mediated differentiation is one of the main themes that was identified from the cross-study analysis. Teacher-led differentiation is clearly effective, especially for guaranteeing access to learning for students of varying aptitude and motivation and so mediating class heterogeneity. Additionally, the findings of this systematic review demonstrate that DI is not limited to contexts with abundant resources. Teachers in big, resource-constrained classes effectively varied instruction using question modification techniques including code-switching and decomposition. These strategies are useful for maintaining learners' interest in the group sessions, even though they place a greater emphasis on access and involvement than individualised worksheets.

Lastly, the findings point to some psychological processes that connect DI to improved attitudes. According to study #6 DI is linked to increased Mathematics Self-Efficacy, learning motivations, which enhances problem-solving abilities. DI lessens the tension that comes with challenging assignments by making sure they are within the student's Zone of Proximal Development (Vygotsky, 1978). Study #3 found qualitative increases in behavioural engagement among young learners using adaptive apps, and study #9 showed significant improvements in confidence and satisfaction.

9.3. Combined Pedagogical and Research Implications

Taken together, the results of the scoping review and the systematic review highlight how closely linguistic structures and teaching methods influence early mathematical development. Phase 1 shows how learning early numeracy can be impacted by certain aspects of language number naming, namely transparency, inversion, and the presence or absence of semantic indicators such singular/plural distinctions. Children's performance on basic mathematical concepts, such as counting, place value, and cardinality, is empirically impacted by such linguistic traits. Phase 2 shows that DI is a successful strategy for improving primary school children's attitudes towards mathematics as well as their academic performance, particularly when combined with adaptive technologies and assessment-driven personalisation. When taken as a whole, these results provide a range of research and pedagogical implications that are pertinent to educators, curriculum designers, and educational researchers.

Language-sensitive mathematics education is desperately needed. Teachers need to be taught to see that student errors, like inversion errors, can be caused by a child's first language's linguistic patterns rather than merely mathematical misunderstandings. For children whose native languages can potentially hinder their early numerical acquisition, early and regular use of DI, especially technology-supported adaptive platforms, could offer crucial supplementary support. Since there is evidence linking language to early numeracy, further longitudinal research is necessary to ascertain if these early linguistic disadvantages continue into more complex mathematics or whether effective instruction eventually eliminates them. In order to develop more pedagogically sound and linguistically responsible instructional models, future research

should also examine the interactions between learners' unique language backgrounds and various DI approaches.

9.3.1. Pedagogical Implications

Given the linguistic diversity present in many contemporary classrooms, these findings underscore the importance of adopting language-sensitive approaches to mathematics instruction. Teachers need to recognise that the number naming structures of children's first languages can shape, and in some cases constrain, the development of early numerical concepts. For example, children whose home languages employ inverted number naming systems are more likely to demonstrate inversion errors when interpreting spoken and written numerals. These findings suggest that such errors should not be interpreted solely as mathematical misconceptions, but rather as linguistically mediated difficulties.

From a pedagogical perspective, this highlights the need for explicit and sustained instruction in place value concepts and the compositional structure of multi-digit numbers. Differentiated instruction can support this process by allowing teachers to provide targeted explanations, visual representations, and structured practice tailored to learners who may be particularly vulnerable to these errors. The use of place value charts, base-ten blocks, and de-compositional strategies may be especially beneficial for learners from linguistically diverse backgrounds.

Similarly, the acquisition of the cardinality principle appears to be influenced by linguistic features such as singular–plural marking. In languages where plural morphology is absent or less salient, children may experience delays in grasping quantity distinctions. To address this, teachers should incorporate differentiated

supports that emphasise quantity through visual modelling, repeated counting activities, and the systematic use of manipulatives. Such approaches can help bridge linguistic gaps by grounding numerical concepts in perceptual and concrete experiences.

Early and sustained implementation of differentiated instruction may therefore serve a compensatory function for children whose language backgrounds place them at risk of slower numerical development. Evidence from studies (Bang et al., 2023; Rodríguez-Martínez et al., 2023) suggests that technology-supported, adaptive DI yields small-to-moderate improvements in mathematics achievement during primary education, with particularly promising effects for younger learners. This implies that early childhood and early primary settings may be especially responsive contexts for the introduction of DI frameworks, particularly in multilingual or multicultural classrooms.

Across both the scoping and systematic reviews, a consistent implication concerns the importance of formative assessment and tailored feedback within DI approaches. Teachers are encouraged to make systematic use of assessment data to identify patterns of error that may be linguistically driven, such as inversion errors or difficulties distinguishing between singular and plural quantities. Differentiated responses to these patterns allow instruction to be more precisely aligned with learners' needs. Emerging evidence suggests that the integration of learning analytics and real-time feedback mechanisms can further enhance the responsiveness of DI by supporting timely instructional adjustments.

However, the effectiveness of DI appears to be closely linked to the quality of its implementation, which in turn depends on teacher expertise. Professional

development is therefore a critical enabling factor. Teachers require support not only in the practical use of DI tools and adaptive platforms, but also in developing an understanding of how linguistic diversity intersects with early numeracy learning. Training that prevents the misattribution of difficulties to general ability and promotes more equitable instructional responses is key.

Finally, while DI emerges as a key pedagogical approach, the findings suggest that it should not be applied in isolation. Rather, DI appears to be most effective when embedded within well-structured, evidence-based instructional practices. For learners with lower prior attainment in particular, DI is likely to be more impactful when combined with explicit instruction, the use of contrasting examples, and scaffolded learning cycles (Ziegler et al., 2021). This integrated approach supports both individual responsiveness and instructional coherence, strengthening the overall effectiveness of mathematics teaching in linguistically diverse classrooms.

Research Focus	Pedagogical Implications / Approaches	Evidence / References
First-language number naming	Recognise that number naming shapes early numerical concepts; interpret inversion errors (e.g., "four-and-twenty") as linguistically mediated.	Scoping review: inversion patterns; Systematic review: Bang et al., 2023; Rodríguez-Martínez et al., 2023
Place value & number composition	Explicit instruction on multi-digit numbers; use DI with visual aids (place value charts, base-ten blocks) and decomposition strategies.	Systematic review: place value interventions reduce errors; Scoping review: gaps in multilingual instruction
Cardinality principle	Support understanding that the final number represents quantity; use visual models, repeated counting, and manipulatives, especially where plural marking is weak.	Scoping review: delays linked to absent plural morphology; Systematic review: effectiveness of visual/manipulative strategies
Early & sustained DI	Introduce DI early to support learners at risk; adaptive, technology-supported DI can improve achievement, especially in multilingual classrooms.	Bang et al., 2023; Rodríguez-Martínez et al., 2023; Systematic review: DI in early primary
Formative assessment & feedback	Identify linguistically driven errors; provide timely, targeted feedback; integrate learning analytics.	Systematic review: formative assessment enhances DI responsiveness; Scoping review: learning analytics emerging
Teacher professional development	Train teachers in DI tools and interpreting linguistic errors; prevent misattribution of language-mediated difficulties.	Systematic review: teacher expertise linked to DI effectiveness; Scoping review: gaps in training
Integration with structured instruction	Embed DI within explicit instruction, contrasting examples, scaffolded cycles; combine individual support with coherent whole-class teaching.	Ziegler et al., 2021; Systematic review: DI more effective when structured
Inclusive & culturally responsive practice	Use multilingual resources, peer support, and culturally relevant examples to enhance engagement and access.	Scoping review: engagement benefits of culturally responsive practice
Ongoing monitoring & adaptation	Regularly review teaching strategies; adapt to performance, cognitive, and linguistic development.	Systematic review: iterative adjustment improves outcomes; Scoping review: need for continuous adaptation

Table 9.1. Pedagogical implications for language-informed mathematics instruction

9.3.2. Research Implications and Future Avenues for Research

From a research perspective, these findings highlight a number of crucial topics that require further exploration. First, longitudinal, cross-linguistic research is required to methodically investigate the ways in which specific linguistic characteristics, such as inversion, transparency, and morphological singular/plural markers, influence mathematical growth over time. Whereas the existing literature suggests important correlations between linguistic number naming conventions and early numerical performance, only longitudinal designs can establish causal links and trace whether early linguistic disadvantages persist or are mitigated by instruction.

Future research is needed to enquire into whether learners' linguistic background affects how successful DI interventions might be. Do DI interventions, for example, work especially well for children whose languages have inverted number names or low transparency? Do some DI approaches, like adaptive and technology-based interventions, work better than others at helping people overcome language-based maths challenges? These are questions that are wholly unexplored, yet.

The preliminary results show improved motivation, self-efficacy, and engagement even if the affective effects of DI are less well proven (Lai et al., 2020; Bang et al., 2023). These could be especially important for students whose language-related difficulties could hinder their learning. Therefore, studies should also look into whether better mathematics attitudes are associated with lower cognitive and linguistic load through differentiation, and whether

these attitude improvements impact achievement advantages.

This offers the opportunity to develop and test adaptive technologies that are specifically designed to identify linguistic error patterns, such as plural confusion, inversion errors, and so on, and provide tailored, personalised feedback. In terms of inclusivity and accuracy, this could significantly enhance early mathematical education. Additionally, more research is required to determine how professional development and teacher preparation affect DI performance, particularly in linguistically varied contexts. The extent to which professional development in language awareness enables teachers to evaluate student errors more effectively and modify instruction accordingly would be worthy of further investigation.

Furthermore, studies should look at how language background, DI methodology, and broader instructional tactics interact. A better understanding of how these factors interact could result in the creation of instructional models that are both pedagogically sound and linguistically responsible.

9.4. Limitations

Notwithstanding the care taken in the design and conduct of this research, several limitations should be acknowledged. This thesis employed a two-stage review approach, comprising an initial scoping review followed by a systematic review. While the scoping review was appropriate for mapping the breadth and characteristics of the literature and informing the focus of the subsequent systematic review, the findings from the scoping review are necessarily descriptive and exploratory rather than evaluative. Whilst scoping reviews are excellent tools for gaining overviews of research areas and identifying gaps,

they are best suited for mapping the breadth of available literature, key concepts, and the types of evidence available. A limitation of scoping reviews is that it is not a method that permits researchers to draw definitive conclusions about the quality of the research or the effectiveness of interventions. Whilst the scoping review reported above carried out an extensive methodical search of the existing literature and rigorously selected a set of studies that directly linked to the research question for the study, its scope is necessarily limited by the very nature of the scoping reviews method. Although comprehensive search strategies were implemented across multiple bibliographic databases, it remains possible that relevant studies were not identified. This may reflect variability in terminology across disciplines, limitations in database indexing, or the exclusion of certain forms of grey literature. The decision to focus on peer-reviewed sources, while supporting methodological transparency, may have contributed to publication bias. In the final step of the scoping review methodology, only a small number of interviews were conducted due to difficulties in recruiting participants. Whilst this is an optional addition to the scoping review design, and whilst the voices of the stakeholders enriched the research process, further participants would have enhanced this stage and could have shed further light on the findings.

The systematic review was limited by heterogeneity across included studies, particularly in relation to study designs, conceptual frameworks, and outcome measures. This variability constrained the extent to which findings could be directly compared or synthesised and precluded the use of quantitative meta-analysis. As a result, synthesis was undertaken narratively, which, although appropriate for the nature of the evidence, necessarily involves a degree of

interpretive judgement. Also, it must be noted that the strength of the conclusions drawn from the systematic review is dependent on the quality of the included primary studies. While formal appraisal was conducted, several studies exhibited methodological limitations, including small sample sizes, varied linguistic and educational contexts, research design, and sample sizes. These factors should be taken into account when interpreting the findings.

9.5. Concluding Remarks

The findings of this study indicate that early mathematical development is shaped, in part, by linguistic characteristics and number naming structures, and suggest that differentiated instruction, particularly when adaptive, personalised, and informed by data, represents a potentially effective approach to addressing learner heterogeneity.

This thesis investigated the intricate connection between early mathematics development and linguistic structures using a two-phase research design. In Phase 1, the scoping review found a clear connection between number-naming conventions and how easily young children acquire numerical concepts. In particular, it highlighted three essential language characteristics: semantic singular/plural markers, transparency, and inversion. More specifically, children who speak more transparent languages, such as Mandarin and Mandarin-based systems, Vietnamese, or Welsh, appear to learn counting and place-value ideas more quickly than children who speak opaque languages, like English or French. However, when students map spoken numbers to written Arabic numbers, they make more inversion errors in languages like German which features extensive inverted patterns. Additionally, it emerged that the presence or lack of semantic

singular/plural markers affected how quickly children speaking languages that do not feature these markers learned the cardinality principle, initially having trouble telling sets of "one" from sets of "many". Taken together, these findings underscore the linguistic dimensions of early mathematical development and carry important pedagogical implications. From a pedagogical perspective, the results highlight the importance of teachers' awareness of the linguistic dimensions of mathematical learning and the use of formative assessment and individualised instructional strategies to support conceptual understanding.

The findings of the systematic review indicated that, as compared to non-differentiated education, differentiated instruction has a positive, significant impact on attitudes toward mathematics and on mathematical achievement. The evidence points to differentiated instruction as a suitable pedagogical approach for addressing the potential student variability brought about by the linguistic and cognitive characteristics described in the earlier chapters of this thesis related to number names and numerical processing. DI was demonstrated to successfully close conceptual gaps in the reviewed empirical data, improving both student attitudes and mathematics achievement. The success of DI frequently depends on accurate diagnosis, which involves using learning analytics or adaptive technologies to pinpoint a student's unique cognitive profile and offer tailored scaffolding. According to the research, DI offers the required mechanism to level the playing field in education, even though linguistic structures might produce conceptual starting places that differ among learners.

At the same time, the evidence also points to the need for further research into differential effects of DI across linguistic groups, the affective outcomes associated with differentiated approaches, and the development of adaptive

technologies and professional development programmes that explicitly attend to the language–mathematics interface. Through continued engagement with these areas, both research and practice may move towards more effective and responsive approaches to early mathematics education in classrooms with linguistic diversity.

References

Abín, A., Núñez, J. C., Rodríguez, C., Cueli, M., García, T., & Rosário, P. (2020). Predicting mathematics achievement in secondary education: The role of cognitive, motivational, and emotional variables. *Frontiers in Psychology, 11*, 876. https://doi.org/10.3389/fpsyg.2020.00876

Ackermans, K., Bakker, M., van Loon, A. M., Kral, M., & Camp, G. (2025). Young learners' motivation, self-regulation and performance in personalized learning. *Computers & Education*, *226*, 105208. https://doi.org/10.1016/j.compedu.2024.105208

Ahmad, M. M. (2024). The pros and cons of systematic review and meta-analysis: Methodological insights and future directions. *Digital Journal of Clinical Medicine, 6*(4) https://doi.org/10.55691/2582-3868.1194

Akhsani, L., Junaedi, I., & Asih, T. S. N. (2025). Exploring insights in the creative thinking process of prospective mathematics teacher students. *Al-Jabar: Jurnal Pendidikan Matematika*, *16*(2), 581-596. https://doi.org/10.24042/ajpm.v16i2.28396

Alber, E. (2012). South Tyrol's education system: Plurilingual answers for monolinguistic spheres? *Europe en formation, 1*, 399-415. https://doi.org/10.3917/eufor.363.0399

Antell, S. E., & Keating, D. P. (1983). Perception of numerical invariance in neonates. *Child Development, 54*(3), 695-701. https://doi.org/10.2307/1130080

Arksey, H., & O'Malley, L. (2005). Scoping studies: Towards a methodological framework.

*International Journal of Social Research Methodology, 8(*1), 19-32. https://doi.org/10.1080/1364557032000119616

Arnold, T. W. (1992). Continuous laying by American Coots in response to

partial clutch removal and total clutch loss. *The Auk, 109*(3), 407-421. https://doi.org/10.2307/4088203

Ashcraft, M. H., & Christy, K. S. (1995). The frequency of arithmetic facts in elementary texts: Addition and multiplication in grades 1–6. *Journal for Research in Mathematics Education, 26*(5), 396-421. https://doi.org/10.5951/jresematheduc.26.5.0396

Aulie, W. (1957). High-layered numerals in Chol (Mayan). *International Journal of American Linguistics, 23*(4), 281-283. https://doi.org/10.1086/463858

Bang, H. J., Li, L., & Flynn, K. (2023). Efficacy of an adaptive game-based math learning app to support personalized learning and improve early elementary school students' learning. *Early Childhood Education Journal, 51*(4), 717-732. https://doi.org/10.1007/s10643-022-01332-3

Banks, W. P., & Coleman, M. J. (1981). Two subjective scales of number. *Perception & Psychophysics, 29*(2), 95-105. https://doi.org/10.3758/BF03205876

Bannister, N. A. (2016). Breaking the spell of differentiated instruction through equity pedagogy and teacher community. Cultural Studies of Science Education, 11(2), 335-347. https://doi.org/10.1007/s11422-016-9766-0

Barrouillet, P., Bernardin, S., & Camos, V. (2004). Time constraints and resource sharing in adults' working memory spans. *Journal of Experimental Psychology: General, 133*(1), 83-100. https://doi.org/10.1037/0096-3445.133.1.83

Baur, S. (2011). Schulsysteme und Lehrerbildung in Südtirol unter dem Aspekt der kulturellen Heterogenität und der Mehrsprachigkeit. In K. Bräu, U. Carle, & I. Kunze (Eds.), *Differenzierung, Integration, Inklusion: Was können wir vom Umgang mit Heterogenität an Kindergärten und Schulen in Südtirol lernen?* (pp. 31-39). Schneider. https://doi.org/10.25162/9783515118965

Beckmann, S. (2002). Mathematics for elementary teachers: Making sense by "explaining why". Department of Mathematics, University of Georgia

Beetsma, D. (2002). Trilingual primary education in Europe. *Fryske Akademy, 93*-106. https://doi.org/10.2307/jj.33169504.15

Belbasis, L., Bellou, V., & Ioannidis, J. P. A. (2022). Conducting umbrella reviews. *BMJ Medicine*, *1*(1), e000071. https://doi.org/10.1136/bmjmed-2021-000071

Bender, A., & Beller, S. (2011). Cultural variation in numeration systems and their mapping onto the mental number line. *Journal of Cross-Cultural Psychology, 42*(4), 579-597. https://doi.org/10.1177/0022022110375346

Beran, M. J., Rumbaugh, D. M., & Savage-Rumbaugh, E. S. (1998). Chimpanzee (Pan troglodytes) counting in a computerized testing paradigm. *The Psychological Record, 48*(1), 3-19. https://doi.org/10.1007/BF03395186

Bijeljac-Babic, R., Bertoncini, J., & Mehler, J. (1993). How do 4-day-old infants categorize multisyllabic utterances? *Developmental Psychology, 29*(4), 711-721. https://doi.org/10.1037/0012-1649.29.4.711

Bloom, P. (2002). *How children learn the meanings of words*. MIT Press.

https://doi.org/10.7551/mitpress/3577.001.0001

Boaler, J., & Brodie, K. (2004, October). The importance, nature and impact of teacher questions. In Proceedings of the twenty-sixth annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education (Vol. 2, pp. 774-782). Chicago, IL: University of Illinois

at Chicago.

Boaler, J. (2016). *Mathematical mindsets: Unleashing students' potential through creative math, inspiring messages and innovative teaching.* Jossey-Bass. https://doi.org/10.1080/14794802.2016.1237374

Boland, A., Cherry, M. G., & Dickson, R. (2017). *Doing a systematic review: A student's guide* (2nd ed.). Sage.

Booth, J. L., & Siegler, R. S. (2006). Developmental and individual differences in pure numerical estimation. *Developmental Psychology, 42*(1), 189-201. https://doi.org/10.1037/0012-1649.42.1.189

Booth, J. L., & Siegler, R. S. (2008). Numerical magnitude representations influence arithmetic learning. *Child Development, 79*(4), 1016-1031. https://doi.org/10.1111/j.1467-8624.2008.01172.x

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2021). *Introduction to meta-analysis* (2nd ed.). John Wiley & Sons. https://doi.org/10.1002/9781119558378

Boysen, S. T. (1993). Counting in chimpanzees: Nonhuman principles and emergent properties of number. In S. T. Boysen & L. A. V. M. O'Hara (Eds.), *The development of numerical competence: Animal and human models* (pp. 39-59). Lawrence Erlbaum

Brannon, E. M., & Terrace, H. S. (1998). Ordering of the numerosities 1 to 9 by monkeys.

Science, 282(5389), 746-749. https://doi.org/10.1126/science.282.5389.746

Brannon, E. M., Wusthoff, C. J., Gallistel, C. R., & Gibbon, J. (2001). Numerical subtraction in the pigeon: Evidence for a linear subjective number scale. *Psychological Science, 12*(3), 238-243. https://doi.org/10.1111/1467-9280.00343

Bryant, P. (2016). Mathematical understanding in the nursery school years. In *Learning and teaching mathematics* (pp. 53-67). Psychology Press.

Butterworth, B. (1999). *What counts: How every brain is hardwired for math*. Free Press.

https://doi.org/10.5860/choice.37-2823

Byeon, S., & Kim, N. (2020). Impact of Korean students' individual learning time on math performance: differential effect of teachers' assessment competency. *Asia Pacific Education Review*, *21*(4), 601-613. https://doi.org/10.1007/s12564-020-09643-z

Campbell, F., Tricco, A. C., Munn, Z., Pollock, D., Saran, A., Sutton, A., & Khalil, H. (2023). Mapping reviews, scoping reviews, and evidence and gap maps (EGMs): The same but different—the "Big Picture" review family. *Systematic Reviews*, *12*(1), 45. https://doi.org/10.1186/s13643-023-02056-1

Carbonneau, K. J., Marley, S. C., & Selig, J. P. (2013). A meta-analysis of the efficacy of teaching mathematics with concrete manipulatives. *Journal of Educational Psychology, 105*(2), 380. https://doi.org/10.1037/a0031084

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

Carmichael, C., Callingham, R., & Watt, H. M. (2017). Classroom motivational environment influences on emotional and cognitive dimensions of student

interest in

mathematics. *ZDM*, *49*(3), 449-460. https://doi.org/10.1007/s11858-016-0831-7

Case, R., & Sowder, J. T. (1990). The development of computational estimation: A neo-Piagetian analysis. *Cognition and Instruction, 7*(2), 79-104. https://doi.org/10.1207/s1532690xci0702_1

Case, R., Okamoto, Y., Griffin, S., McKeough, A., Bleiker, C., Henderson, B., & Keating, D.

P. (1996). The role of central conceptual structures in the development of children's thought. *Monographs of the Society for Research in Child Development, 61*(4), i-295. https://doi.org/10.2307/1166094

Cazden, C. B. (1968). The acquisition of noun and verb inflections. *Child Development, 39*(2), 433-448. https://doi.org/10.2307/1127375

Chapin, S. H., O'Connor, M. C., & Anderson, N. C. (2009). *Classroom discussions: Using math talk to help students learn, Grades K-6.* Math Solutions.

Choi, G. J., & Kang, H. (2023). Introduction to umbrella reviews as a useful evidence-based practice. *Journal of Lipid and Atherosclerosis, 12*(1), 3–11. https://doi.org/10.12997/jla.2023.12.1.3

Chrisomalis, S. (2004). A cognitive typology for numerical notation. *Cambridge Archaeological Journal, 14*(1), 37-51. https://doi.org/10.1017/S0959774304000021

Church, R. M., & Meck, W. H. (1984). The numerical attribute of stimuli. In H. L. Roitblat,

T. G. Bever, & H. Cohen (Eds.), *Animal cognition: A tribute to Donald A. R. at his 80th birthday* (pp. 141-161). Lawrence Erlbaum.

https://doi.org/10.1016/0140-1750(86)90020-5

Condry, K. F., & Spelke, E. S. (2008). The development of language and abstract concepts: The case of natural number. *Journal of Experimental Psychology: General, 137*(1), 22-38. https://doi.org/10.1037/0096-3445.137.1.22

Cook, D. J., Mulrow, C. D., & Haynes, R. B. (1997). Systematic reviews: synthesis of best evidence for clinical decisions. *Annals of Internal Medicine, 126*(5), 376-380. https://doi.org/10.7326/0003-4819-126-5-199703010-00006

Coupe, A. R. (2012). Overcounting numeral systems and their relevance to subgrouping in the Tibeto-Burman languages of Nagaland. *Language and Linguistics, 13*(1), 193-220. https://doi.org/10.1017/S1603052212000049

Covidence Systematic Review Software (2024). Veritas Health Innovation, Melbourne, Australia. Retrieved 18 December, 2024 from www.covidence.org

Creswell, J. W., & Creswell, J. D. (2018). Research design: Qualitative, quantitative, and mixed methods approaches (5th ed.). Sage.

Dansk Skoleforening for Sydslesvig e.V. (2021, December 13). *Geschichte*. Retrieved December 4, 2024 from http://www.skoleforeningen.org/foreningen/historie?showgerman=1

Davis, G. W. (1998). *Germanische Zahlwörter: Sprach-und kulturgeschichtliche Untersuchungen insbesondere zur Zahl 12.* Akademie Verlag. https://doi.org/10.1006/hmat.1999.2222

Davis, H. (1984). Discrimination of the number three by a raccoon (Procyon lotor). *Animal Learning & Behavior, 12*(4), 409-413. https://doi.org/10.3758/BF03200917

Davis, H., & Bradford, S. A. (1986). Counting behavior by rats in a simulated natural environment. *Ethology, 73*(4), 265-280. https://doi.org/10.1111/j.1439-0310.1986.tb00692.x

Davis, H., & Bradford, S. A. (1991). Numerically restricted food intake in the rat in a free-feeding situation. *Animal Learning & Behavior, 19*(3), 215-222. https://doi.org/10.3758/BF03198685

Davis, H., & Pérusse, R. (1988). Numerical competence in animals: Definitional issues, current evidence, and a new research agenda. *Behavioral and Brain Sciences, 11*(4), 561-579. https://doi.org/10.1017/S0140525X00057946

de Jong, E. J. (2011). *Foundations for multilingualism in education: From principles to practice.* Caslon Publishing. https://doi.org/10.1007/s10993-012-9256-5

Dehaene, S. (2011). *The number sense: How the mind creates mathematics.* Oxford University Press USA. https://doi.org/10.1080/00029890.1998.12004997

Dehaene, S., & Cohen, L. (1994). Dissociable mechanisms of subitizing and counting: Neuropsychological evidence from simultanagnosic patients. *Journal of Experimental Psychology: Human Perception and Performance, 20*(5), 958-975. https://doi.org/10.1037/0096-1523.20.5.958

Dehaene, S., & Marques, J. F. (2002). Cognitive neuroscience: Scalar variability in price estimation and the cognitive consequences of switching to the euro. *The Quarterly Journal of Experimental Psychology: Section A, 55*(3), 705-731. https://doi.org/10.1080/02724980143000589

Dehaene, S., Bossini, S., & Giraux, P. (1993). The mental representation of

parity and number magnitude. *Journal of Experimental Psychology: General, 122*(3), 371-396. https://doi.org/10.1037/0096-3445.122.3.371

Dehaene, S., Dupoux, E., & Mehler, J. (1990). Is numerical comparison digital? Analogical and symbolic effects in two-digit number comparison. *Journal of Experimental Psychology: Human Perception and Performance, 16*(3), 626-641. https://doi.org/10.1037/0096-1523.16.3.626

Dehaene, S., Izard, V., Spelke, E., & Pica, P. (2008). Log or linear? Distinct intuitions of the number scale in Western and Amazonian indigenous cultures. *Science, 320*(5880), 1217-1220. https://doi.org/10.1126/science.1156548

Dehaene, S., Molko, N., Cohen, L., & Wilson, A. J. (2004). Arithmetic and the brain. *Current Opinion in Neurobiology, 14*(2), 218-224. https://doi.org/10.1016/j.conb.2004.03.008

Denzin, N. K., & Lincoln, Y. S. (Eds.). (2018). *The Sage handbook of qualitative research* (5th ed.). Sage.

Dowker, A., & Li, A. M. (2019). English and Chinese children's performance on numerical tasks. *Frontiers in Psychology, 9*, 2731. https://doi.org/10.3389/fpsyg.2018.02731

Dowker, A., & Roberts, M. (2015). Does the transparency of the counting system affect children's numerical abilities? *Frontiers in Psychology, 6,* 945. https://doi.org/10.3389/fpsyg.2015.00945

Dowker, A., Bala, S., & Lloyd, D. (2008). Linguistic influences on mathematical development: How important is the transparency of the counting system? *Philosophical Psychology, 21*(4), 523-538. https://doi.org/10.1080/09515080802244312

Durkin, K., Shire, B., Riem, R., Crowther, R. D., & Rutter, D. R. (1986). The social and linguistic context of early number word use. British Journal of Developmental Psychology, 4(3), 269-288. https://doi.org/10.1111/j.2044-835X.1986.tb01013.x

Duval, R. (2006). A cognitive analysis of problems of comprehension in the learning of mathematics. *Educational Studies in Mathematics, 61*(1), 103-131. https://doi.org/10.1007/s10649-006-0400-2

Ebersbach, M., Luwel, K., Frick, A., Onghena, P., & Verschaffel, L. (2008). The relationship between the shape of the mental number line and familiarity with numbers in 5- to 9-year-old children: Evidence for a segmented linear model. *Journal of Experimental Child Psychology, 99*(1), 1-17. https://doi.org/10.1016/j.jecp.2007.08.004

Edgeworth, M., & Edgeworth, R. L. (1798). *Practical education*. J. Johnson. https://doi.org/10.1017/cbo9781139207645

Ejersbo, L. R., & Misfeldt, M. (2011, February). Danish number names and number concepts. In *Proceedings of the 7th Conference of European Research on Mathematics Education (CERME 7)*. University of Rzeszów.

Essien, A. A. (2018). The role of language in the teaching and learning of early grade mathematics: An 11-year account of research in Kenya, Malawi and South Africa. *African Journal of Research in Mathematics, Science and Technology Education, 22*(1), 48-59. https://doi.org/10.1080/18117295.2018.1434453

Febrilia, Y., Mardati, A., & Machbub, A. D. Z. (2025). Effects of differentiated instruction on primary students' mathematics achievement: A quasi-experimental study. *Al-Jabar: Jurnal Pendidikan Matematika, 16*(2).

https://doi.org/10.24042/ajpm.v16i2.27541

Feigenson, L., & Carey, S. (2003). Tracking individuals via object-files: Evidence from infants' manual search. *Developmental Science, 6*(5), 568-584. https://doi.org/10.1111/1467-7687.00303

Feigenson, L., & Carey, S. (2005). On the limits of infants' quantification of small object arrays. *Cognition, 97*(3), 295-313. https://doi.org/10.1016/j.cognition.2005.01.009

Feigenson, L., Carey, S., & Hauser, M. (2002). The representations underlying infants' choice of more: Object files versus analog magnitudes. *Psychological Science, 13*(2), 150-156. https://doi.org/10.1111/1467-9280.00427

Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *Trends in Cognitive Sciences, 8*(7), 307-314. https://doi.org/10.1016/j.tics.2004.05.002

Fernandez, R. S., Sharifnia, A. M., & Khalil, H. (2025). Umbrella reviews: A methodological guide. *European Journal of Cardiovascular Nursing, 24*(6), 996–1002. https://doi.org/10.1093/eurjcn/zvaf012

Fischer, J.-P. (1992). Subitising: The discontinuity after three. (C. Greenbaum, Trans.). In J. Bideaud, C. Meljac, & J.-P. Fischer (Eds.), *Pathways to number: Children's developing numerical abilities* (pp. 191–208). Lawrence Erlbaum Associates.

Fosnot, C. T. (2013). *Constructivism: Theory, perspectives, and practice*. Teachers College Press.

Freeman, N. H., Antonucci, C., & Lewis, C. (2000). Representation of the cardinality principle: Early conception of error in a counterfactual test. *Cognition*, 74(1), 71-89. https://doi.org/10.1016/S0010-0277(00)00030-6

Frege, G. (1980). *Die Grundlagen der Arithmetik* [The foundations of arithmetic] (J. L. Austin, Trans.). Blackwell. (Original work published 1884)

https://doi.org/10.1007/978-3-658-49495-7_9

Frye, D., Braisby, N., Lowe, J., Maroudas, C., & Nicholls, J. (1989). Young children's understanding of counting and cardinality. *Child Development, 60*(5), 1158-1171. https://doi.org/10.2307/1130806

Fuson, K. C. (1988). Children's counting and concepts of number. Springer-Verlag.

https://doi.org/10.1007/978-1-4612-3754-9

Fuson, K. C., & Kwon, Y. (1992). Korean children's understanding of multidigit addition and subtraction. *Child Development, 63*(2), 491-506. https://doi.org/10.2307/1131226

Fuson, K. C., & Li, Y. (2009). Cross-cultural issues in linguistic, visual-quantitative, and written-numeric supports for mathematical thinking. *ZDM*, *41*(6), 793-808. https://doi.org/10.1007/s11858-009-0229-5

Fuson, K. C., Richards, J., & Briars, D. J. (1982). The acquisition and elaboration of the number word sequence. In *Children's logical and mathematical cognition* (pp. 33-92). Springer. https://doi.org/10.1007/978-1-4613-9466-2_2

Ganann, R., Ciliska, D., & Thomas, H. (2010). Expediting systematic reviews: Methods and implications of rapid reviews. *Implementation Science, 5*(1), 56. https://doi.org/10.1186/1748-5908-5-56

Gandara, F., & Randall, J. (2019). Assessing mathematics proficiency of multilingual students: The case for translanguaging in the Democratic Republic of the Congo. *Comparative Education Review*, *63*(1), 58-78. https://doi.org/10.1086/701065

García, O., & Wei, L. (2014). *Translanguaging: Language, bilingualism and education.* Palgrave Macmillan. https://doi.org/10.46538/hlj.11.3.4

Garritty, C., Nussbaumer-Streit, B., Hamel, C., & Devane, D. (2024). Rapid reviews methods series: Assessing the appropriateness of conducting a rapid review. *BMJ Evidence-Based Medicine, 30*(1), 55–60. https://doi.org/10.1136/bmjebm-2023-112722

Gay, G. (2010). *Culturally responsive teaching: Theory, research, and practice* (2nd ed.). Teachers College Press. https://doi.org/10.31046/wabashcenter.v1i3.1798

Geary, D. C., Bow-Thomas, C. C., Liu, F., & Siegler, R. S. (1996). Development of arithmetical competencies in Chinese and American children: Influence of age, language, and schooling. *Child Development, 67*(5), 2022-2044. https://doi.org/10.2307/1131600

Gelman, R., & Gallistel, C. R. (1986). *The child's understanding of number*. Harvard University Press. https://doi.org/10.4159/9780674037533

Gibbon, J., & Church, R. M. (1981). Time left: Linear versus logarithmic subjective time.

Journal of Experimental Psychology: Animal Behavior Processes, 7(2), 87-108. https://doi.org/10.1037/0097-7403.7.2.87

Gibbons, P. (2002). *Scaffolding language, scaffolding learning: Teaching second language learners in the mainstream classroom.* Heinemann.

Gibson, D. J., & Shusterman, A. (2009). Evidence for very early understanding that "two" means "a pair." *Poster presented at the biennial meeting of the Society for Child Development, Denver, CO.* https://doi.org/10.1037/e516472011-007

Goddard, L. (1961). Counting. *Australasian Journal of Philosophy, 39*(3), 223-240. https://doi.org/10.1080/00048406185515O1

Guerrero, D., Hwang, J., Boutin, B., Roeper, T., & Park, J. (2020). Is thirty-two three

tens and two ones? The embedded structure of cardinal numbers. *Cognition, 203*, 104331. https://doi.org/10.1016/j.cognition.2020.104331

Gurevitch, J., Koricheva, J., Nakagawa, S., & Stewart, G. (2018). Meta-analysis and the science of research synthesis. *Nature, 555*(7695), 175–182. https://doi.org/10.1038/nature25753

Gutstein, E. (2006). *Reading and writing the world with mathematics: Toward a pedagogy for social justice.* Routledge. https://doi.org/10.1080/10986060709336607

Guyatt, G. H., Sackett, D. L., Sinclair, J. C., Hayward, R., Cook, D. J., & Cook, R. J. (1995).

Users' guides to the medical literature: IX. A method for grading health care recommendations. *JAMA, 274*(22), 1800–1804. https://doi.org/10.1001/jama.1995.03530220066035

Hall, T. (2002). *Differentiated Instruction.* Effective Classroom Practices Report, National Center on Accessing the General Curriculum, Office of Special Education Programs, U.S. Department of Education.

Hamann, M. S., & Ashcraft, M. H. (1986). Textbook presentations of the basic addition facts. *Cognition and Instruction, 3*(3), 173-202. https://doi.org/10.1207/s1532690xci0303_1

Hanin, V., & Van Nieuwenhoven, C. (2016). The influence of motivational and emotional factors in mathematical learning in secondary education. *European Review of Applied Psychology, 66*(3), 127-138. https://doi.org/10.1016/j.erap.2016.04.006

Hauser, M. D., MacNeilage, P., & Ware, M. (1996). Numerical representations in primates.

Proceedings of the National Academy of Sciences, 93(4), 1514-1517. https://doi.org/10.1073/pnas.93.4.1514

Hecox, K. E., & Hagen, J. W. (1971). Estimates and estimate-based inferences in young children. *Journal of Experimental Child Psychology, 11*(1), 106-123. https://doi.org/10.1016/0022-0965(71)90034-0

Helmreich, I., Zuber, J., Pixner, S., Kaufmann, L., Nuerk, H. C., & Moeller, K. (2011). Language effects on children's nonverbal number line estimations. *Journal of Cross-Cultural Psychology, 42*(4), 598-613. https://doi.org/10.1177/0022022110365534

Higgins, J. P., & Green, S. (Eds.). (2008). *Cochrane handbook for systematic reviews of interventions.* https://doi.org/10.1002/9780470712184

Hickey, R. (2011). *The dialects of Irish: study of a changing landscape* (Vol. 230). Walter de Gruyter. https://doi.org/10.1515/9783110238303

Hitch, G., Cundick, J., Haughey, M., Pugh, R., & Wright, H. (1987). Aspects of counting in children's arithmetic. In J. A. Sloboda & D. Rogers (Eds.), *Cognitive processes in mathematics* (pp. 26–41). Oxford University Press.

Ho, C. S. H., & Fuson, K. C. (1998). Children's knowledge of teen quantities as tens and ones: Comparisons of Chinese, British, and American kindergartners. *Journal of Educational Psychology, 90*(3), 536. https://doi.org/10.1037/0022-0663.90.3.536

Howe, R. (2019). Learning and using our base-10 place value number system: Theoretical perspectives and twenty-first century uses. *ZDM, 51*(1), 57-68. https://doi.org/10.1007/s11858-018-0987-5

Huedo-Medina, T. B., Sánchez-Meca, J., Marín-Martínez, F., & Botella, J. (2006). Assessing heterogeneity in meta-analysis: Q statistic or I2 index? *Psychological Methods, 11*(2), 193–206. https://doi.org/10.1037/1082-989X.11.2.193

Hunt, H., Pollock, A., Campbell, P., Estcourt, L., & Brunton, G. (2018). An introduction to overviews of reviews: Planning a relevant research question and objective for an overview. *Systematic Reviews, 7*(1), 1-9.

https://doi.org/10.1186/s13643-018-0757-2

Huntley-Fenner, G. (2001). Children's understanding of number is similar to adults' and rats': Numerical estimation by 5–7-year-olds. *Cognition, 78*(3), B27-B40. https://doi.org/10.1016/S0010-0277(00)00158-1

Hurford, J. R. (2011). *The linguistic theory of numerals* (Vol. 16). Cambridge University Press.

Huttenlocher, J., & Smiley, P. (1987). Early word meanings: The case of object names. *Cognitive Psychology, 19*(1), 63-89. https://doi.org/10.1016/0010-0285(87)90003-2

Irvine, A., Gabbay, D. M., Thagard, P., & Woods, J. (2009). *Philosophy of mathematics.* Elsevier. https://doi.org/10.1016/b978-0-444-51555-1.50002-x

Ito, Y., & Hatta, T. (2004). Spatial structure of quantitative representation of numbers: Evidence from the SNARC effect. *Memory & Cognition, 32*(4), 662-673. https://doi.org/10.3758/BF03195863

JBI (2015). *JBI manual for evidence synthesis*. Retrieved December 10, 2024 from https://jbi.global/scoping-review-network/resources

Jensen, E. M., Reese, E. P., & Reese, T. W. (1950). The subitising and counting of visually presented fields of dots. *Journal of Psychology, 30*(2), 363-392. https://doi.org/10.1080/00223980.1950.9916251

Jing, L. (1984). A study of development of number notion in young children. *Psychological Science, 3,* 115-119.

Jones, G. A., Thornton, C. A., Putt, I. J., Hill, K. M., Mogill, A. T., Rich, B. S., & Van Zoest, L.R. (1996). Multidigit number sense: A framework for instruction and assessment. *Journal for Research in Mathematics Education, 27*(3), 310-336. https://doi.org/10.2307/749809

Jourdain, L. V., & Sharma, S. (2016). Language challenges in mathematics education: A literature review. *Waikato Journal of Education, 21*(2).

https://doi.org/10.15663/wje.v21i2.269

Khalil, H., & Tricco, A. C. (2022). Differentiating between mapping reviews and scoping reviews in the evidence synthesis ecosystem. *Journal of Clinical Epidemiology, 149,* 175–182. https://doi.org/10.1016/j.jclinepi.2022.05.012

Kamii, C., & Kysh, J. (2006). The difficulty of "length × width": Is a square the unit of measurement? *Journal of Mathematical Behavior, 25*(2), 105-115. https://doi.org/10.1016/j.jmathb.2006.01.001

Kamii, C., Rummelsburg, J., & Kari, A. (2005). Teaching arithmetic to low-performing, low-SES first graders. *Journal of Mathematical Behavior, 24*(1), 39-50. https://doi.org/10.1016/j.jmathb.2005.01.002

Kaufman, E. L., Lord, M. W., Reese, T. W., & Volkmann, J. (1949). The discrimination of visual number. *The American Journal of Psychology, 62*(4), 498-525. https://doi.org/10.2307/1418682

Kazemi, E., & Stipek, D. (2001). Promoting conceptual thinking in four upper-elementary mathematics classrooms. *Elementary School Journal, 102*(1), 59–80. https://doi.org/10.1086/499693

Kim, C., Park, S. W., & Cozart, J. (2014). Affective and motivational factors of learning in online mathematics courses. *British Journal of Educational Technology, 45*(1), 171-185. https://doi.org/10.1111/j.1467-8535.2012.01382.x

Koechlin, E., Dehaene, S., & Mehler, J. (1997). Numerical transformations in five-month-old human infants. *Mathematical Cognition, 3*(2), 89-104. https://doi.org/10.1080/135467997387508

Koehler, O. (1941). Vom Erlernen unbenannter Anzahlen bei Vögeln. *Naturwissenschaften, 29*(14-15), 201-218. https://doi.org/10.1007/BF01494072

Koehler, O. (1943). "Zähl"-Versuche an einem Kolkraben und Vergleichsversuche an Menschen. *Zeitschrift für Tierpsychologie, 5*(3), 575-712. https://doi.org/10.1111/j.1439-0310.1943.tb01669.x

Koehler, O., & Wachholtz, R. (1936). Weitere Versuche an der Taube Nichtweiß zur Frage des Zählvermögens. *Verhandlungen der Deutschen Zoologischen Gesellschaft, 38*, 211-236.

Koehler, O., & Wachholtz, R. (1950, June). Vögel erlernen unbenannte Anzahlen. In *Proceedings of the 9th International Congress of Psychology.*

Konig, E., & Van der Auwera, J. (2013). *The Germanic languages*. Routledge https://doi.org/10.4324/9781315812786

Kutter, E. F., Dehnen, G., Borger, V., Surges, R., Mormann, F., & Nieder, A. (2023). Distinct neuronal representation of small and large numbers in the human medial temporal lobe. *Nature Human Behaviour, 7*(11), 1998-2007. https://doi.org/10.1038/s41562-023-01585-0

Kyllonen, P. C., Lohman, D. F., & Woltz, D. J. (1984). Componential modeling of alternative strategies for performing spatial tasks. *Journal of Educational Psychology, 76*(6), 1325-1345. https://doi.org/10.1037/0022-0663.76.6.1325

Labinowicz, E. (1980). *The Piaget primer: Thinking, learning, teaching.* Addison-Wesley.

Lam, L. Y., & Ang, T. S. (2004). Fleeting footsteps: Tracing the conception of arithmetic and algebra in ancient China. *World Scientific*. https://doi.org/10.1142/9789812567253

Lai, C. P., Zhang, W., & Chang, Y. L. (2020). Differentiated instruction enhances sixth-grade students' mathematics self-efficacy, learning motives, and problem-solving skills. *Social Behavior and Personality, 48*(6), 1-14. https://doi.org/10.2224/sbp.9094

Laski, E. V., & Yu, Q. (2014). Number line estimation and mental addition: Examining the potential roles of language and education. *Journal of Experimental Child Psychology, 117,* 29-44. https://doi.org/10.1016/j.jecp.2013.09.008

Lavrijsen, J., Dockx, J., Struyf, E., & Verschueren, K. (2022). Class composition, student achievement, and the role of the learning environment. *Journal of Educational Psychology, 114*(3), 498. https://doi.org/10.1037/edu0000709

Le Corre, M., & Carey, S. (2007). One, two, three, four, nothing more: An investigation of the conceptual sources of the verbal counting principles. *Cognition, 105*(2), 395-438. https://doi.org/10.1016/j.cognition.2006.10.011

Le Corre, M., Li, P., & Jia, G. (2003, April). On the role of singular/plural in number word learning. Poster presented at the Biennial Meeting of the Society for Research in Child Development, Tampa, FL.

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. https://doi.org/10.1016/j.cogpsych.2005.11.002

Lean, G. A. (1992). Counting systems of Papua New Guinea and Oceania. Papua New Guinea University of Technology.

Lee, M. D., & Sarnecka, B. W. (2010). A model of knower-level behavior in number concept development. *Cognitive Science, 34*(1), 51-67. https://doi.org/10.1111/j.1551-6709.2009.01059.x

LeFevre, J. A., Greenham, S. L., & Waheed, N. (1993). The development of procedural and conceptual knowledge in computational estimation. *Cognition and Instruction, 11*(2), 95-132. https://doi.org/10.1207/s1532690xci1102_1

Levac, J., Colquhoun, H., & O'Brien, K. K. (2010). Scoping studies: Advancing the methodology. *Implementation Science, 5*(1), 69. https://doi.org/10.1186/1748-5908-5-69

Levey, S. (2023). Universal design for learning. *Journal of Education, 203*(2), 479-487. https://doi.org/10.1177/00220574211031954

Li, P., Le Corre, M., Shui, R., Jia, G., & Carey, S. (2003, October). Effects of plural

syntax on number word learning: A cross-linguistic study. In *Proceedings of the 28th Boston University Conference on Language Development*, Boston, MA. https://doi.org/10.1037/e625392011-001

Lieberman, E., Michel, J. B., Jackson, J., Tang, T., & Nowak, M. A. (2007). Quantifying the evolutionary dynamics of language. *Nature, 449*(7163), 713-716. https://doi.org/10.1038/nature06134

Link, T., Huber, S., Nuerk, H. C., & Moeller, K. (2014). Unbounding the mental number line—New evidence on children's spatial representation of numbers. *Frontiers in Psychology, 4*, 1021. https://doi.org/10.3389/fpsyg.2013.01021

Lyon, B. E. (2003). Egg recognition and counting reduce costs of avian conspecific brood parasitism. *Nature, 422*(6931), 495-499. https://doi.org/10.1038/nature01557

Mac Mathúna, S. (2006). Vigesimal Numerals in Irish and Celtic. In J. Clancy (Ed.), *Celtic Language and Literature Studies* (pp. 77–94). Dublin: Four Courts Press.

Macaro, E. (2019). Systematic reviews in applied linguistics. In J. McKinley & H. Rose (Eds.), *The Routledge handbook of research methods in applied linguistics* (pp. 230–239). Routledge.

Magargee, S. D., & Beauford, J. E. (2016). Do explicit number names accelerate pre-kindergarteners' numeracy and place value acquisition? *Educational Studies in Mathematics, 92*(2), 179-192. https://doi.org/10.1007/s10649-016-9693-6

Mandler, G., & Shebo, B. J. (1982). Subitising: An analysis of its component processes. *Journal of Experimental Psychology: General, 111*(1), 1-22. https://doi.org/10.1037/0096-3445.111.1.1

Marshall, K. (2016). Rethinking differentiation—Using teachers' time most effectively. *Phi Delta Kappan, 98*(1), 8-13. https://doi.org/10.1177/0031721716666046

Maslow, A. (1965). *Self-actualization and beyond*. Van Nostrand.

Matsuzawa, T. (1985). Use of numbers by a chimpanzee. *Nature, 315*(6014), 57-59. https://doi.org/10.1038/315057a0

Maulana, R., Smale-Jacobse, A., Helms-Lorenz, M., Chun, S., & Lee, O. (2020). Measuring differentiated instruction in The Netherlands and South Korea: factor structure equivalence, correlates, and complexity level. *European Journal of Psychology of Education, 35*(4), 881-909. https://doi.org/10.1007/s10212-019-00446-4

Mazaudon, M. (2009). North East Indian linguistics: Number-building in Tibeto-Burman languages. In *NEILS 2 (North-East India Languages Symposium)*, Gauhati, Assam, India, 5–9 February 2007. Retrieved December 9, 2024, from https://shs.hal.science/file/index/docid/273445/filename/Number_building_NEILS2.pdf

McCloskey, M. (1992). Cognitive mechanisms in numerical processing: Evidence from acquired dyscalculia. *Cognition, 44*(1-2), 107-157. Retrieved December 9, 2024 from https://doi.org/10.1016/0010-0277(92)90063-T

McCloskey, M., Caramazza, A., & Basili, A. (1985). Cognitive mechanisms in number processing and calculation: Evidence from dyscalculia. *Brain and Cognition, 4*(2), 171-196. https://doi.org/10.1016/0278-2626(85)90017-9

McComb, K., Packer, C., & Pusey, A. (1994). Roaring and numerical assessment in contests between groups of female lions, Panthera leo. *Animal Behaviour, 47*(2), 379-387. https://doi.org/10.1006/anbe.1994.1039

McDonald, S. (2020). Fit for the challenge: The new Cochrane handbook. *Journal of Health Information and Libraries Australasia, 1*(2), 5-10. https://doi.org/10.55999/johila.v1i2.29

Mechner, F. (1958). Probability relations within response sequences under ratio reinforcement. *Journal of the Experimental Analysis of Behavior, 1*(2), 109-116. https://doi.org/10.1901/jeab.1958.1-109

Mechner, F., & Guevrekian, L. (1962). Effects of deprivation upon counting and timing in rats. *Journal of the Experimental Analysis of Behavior, 5*(4), 463-466. https://doi.org/10.1901/jeab.1962.5-463

Menninger, K. (1979). *Zahlwort und Ziffer: Eine Kulturgeschichte der Zahl* (Vol. 1). Vandenhoeck & Ruprecht. https://doi.org/10.1002/zamm.19590390725

Menninger, K. (2013). *Number words and number symbols: A cultural history of numbers.* Courier Corporation. https://doi.org/10.1017/s0022226700013232

Mervis, C. B., & Johnson, K. E. (1991). Acquisition of the plural morpheme: A case study. *Developmental Psychology, 27*(2), 222-230. https://doi.org/10.1037/0012-1649.27.2.222

Meyer, A., Rose, D. H., & Gordon, D. (2014). Universal design for learning: Theory and practice. *CAST Professional Publishing*. https://doi.org/10.4135/9781446282236.n30

Miller, K. F., & Stigler, J. W. (1987). Counting in Chinese: Cultural variation in a basic cognitive skill. *Cognitive Psychology, 19*(3), 297-325. https://doi.org/10.1016/0010-0285(87)90013-6

Miller, K. F., & Zhu, J. (1991). The trouble with teens: Accessing the structure of number names. *Journal of Memory and Language, 30*(1), 48-68. https://doi.org/10.1016/0749-596X(91)90006-M

Miller, K. F., Smith, C. M., Zhu, J., & Zhang, H. (1995). Preschool origins of cross-national differences in mathematical competence: The role of number-naming systems. *Psychological Science, 6*(1), 56-60. https://doi.org/10.1111/j.1467-9280.1995.tb00340.x

Mitchell, R. W., Yao, P., Sherman, P. T., & O'Regan, M. (1985). Discriminative responding of a dolphin (Tursiops truncatus) to differentially rewarded stimuli. *Journal of Comparative Psychology, 99*(2), 218-228.

https://doi.org/10.1037//0735-7036.99.2.218

Miura, I. T., & Okamoto, Y. (1989). Comparisons of US and Japanese first graders' cognitive representation of number and understanding of place value. *Journal of Educational Psychology, 81*(1), 109. https://doi.org/10.1037/0022-0663.81.1.109

Miura, I. T., & Okamoto, Y. (2013). Language supports for mathematics understanding and performance. In *The development of arithmetic concepts and skills* (pp. 251–264). Routledge. https://doi.org/10.4324/9781410607218-13

Miura, I. T., Kim, C. C., Chang, C. M., & Okamoto, Y. (1988). Effects of language characteristics on children's cognitive representation of number: Cross-national comparisons. *Child Development, 59*(6), 1445–1450. https://doi.org/10.2307/1130659

Miura, I. T., Okamoto, Y., Kim, C. C., Steere, M., & Fayol, M. (1993). First graders' cognitive representation of number and understanding of place value: Cross-national comparisons: France, Japan, Korea, Sweden, and the United States. *Journal of Educational Psychology, 85*(1), 24. https://doi.org/10.1037/0022-0663.85.1.24

Miura, I. T., Okamoto, Y., Kim, C. C., Chang, C. M., Steere, M., & Fayol, M. (1994). Comparisons of children's cognitive representation of number: China, France, Japan, Korea, Sweden, and the United States. *International Journal of Behavioral Development, 17*(3), 401-411. https://doi.org/10.1017/S014271640000713X

Moeller, K., Pixner, S., Kaufmann, L., & Nuerk, H. C. (2009). Children's early mental number line: Logarithmic or decomposed linear? *Journal of Experimental Child Psychology, 103*(4), 503–515. https://doi.org/10.1016/j.jecp.2009.02.006

Moher, D., Stewart, L., & Shekelle, P. (2015). All in the family: Systematic reviews,

rapid reviews, scoping reviews, realist reviews, and more. *Systematic Reviews, 4*(1), 1–2. https://doi.org/10.1186/s13643-015-0163-7

Mohseni, M., Ameri, H., & Arab-Zozani, M. (2022). Potential limitations in systematic review studies assessing the effect of the main intervention for treatment/therapy of COVID-19 patients: An overview. *Frontiers in Medicine, 9*, 966632. https://doi.org/10.3389/fmed.2022.966632

Moons, P., Goossens, E., & Thompson, D. R. (2021). Rapid reviews: The pros and cons of an accelerated review process. *European Journal of Cardiovascular Nursing, 20*(5), 515–519. https://doi.org/10.1093/eurjcn/zvab041

Morin, J. E., & Franks, D. J. (2009). Why do some children have difficulty learning mathematics? Looking at language for answers. Preventing School Failure: *Alternative Education for Children and Youth, 54*(2), 111–118. https://doi.org/10.1080/10459880903217861

Moschkovich, J. N. (2007). Bilingual mathematics learners: How views of language, bilingual learners, and mathematical communication impact instruction. In N. Nasir & P. Cobb (Eds.), *Diversity, equity, and access to mathematical ideas* (pp. 89–104). Teachers College Press.

Moschkovich, J. N. (2010). Language (s) and learning mathematics: Resources, challenges, and issues for research. Munn, Z., Peters, M. D., Stern, C., Tufanaru, C., McArthur, A., & Aromataris, E. (2018). Systematic review or scoping review? Guidance on selecting the review type for evidence synthesis. *BMC Medical Research Methodology, 18*(1), 1–7. https://doi.org/10.1186/s12874-018-0611-x

Moschkovich, J. N. (2011). Supporting mathematical reasoning and sense making for English learners. In M. E. Strutchens & J. R. Quander (Eds.), *Focus in high school mathematics: Fostering reasoning and sense making for all students* (pp. 17–36). National Council of Teachers of Mathematics.

Muthomi, M. W., & Mbugua, Z. K. (2014). Effectiveness of differentiated instruction on secondary school students achievement in mathematics. *International Journal of Applied Science and Technology, 4*(1), 116-128. https://ijast.thebrpi.org/journals/Vol_4_No_1_January_2014/12.pdf

Navaitienė, J., & Stasiūnaitienė, E. (2021). The goal of the universal design for learning: Development of all to expert learners. In L. Augustine, M. H. Shabestari, & E. Siljehag (Eds.), *Improving inclusive education through universal design for learning* (pp. 23–57). Springer International Publishing. https://doi.org/10.1007/978-3-030-80658-3_2

Negen, J., & Sarnecka, B. W. (2012). Number-concept acquisition and general vocabulary development. *Child Development, 83*(6), 2019–2027. https://doi.org/10.1111/j.1467-8624.2012.01815.x

Ní Chiosáin, M., & Tallerman, M. (2011). Understanding Syntax in Irish. *Linguistics Today, 89*(2), 145–172. https://doi.org/10.4324/9780203778579

Nikoloska, A. (2009). Development of the cardinality principle in Macedonian preschool children. *Psihologija, 42*(4), 459–475.

Ó Baoill, D. P. (2009). Irish. In M. J. Ball & J. Fife (Eds.), *The Celtic languages* (pp. 177–243). Routledge. https://doi.org/10.1017/cbo9780511551871.003

Ó Murchú, M. (1985). The standardisation of Irish numerals: an overview. *Ériu, 36*, 120–139.

Ó Siadhail, M. (1989). *Modern Irish: grammatical structure and dialectal variation.* Cambridge University Press https://doi.org/10.1016/0024-3841(91)90037-6

Okamoto, Y., & Case, R. (1996). The role of central conceptual structures in the development of children's thought: II. Exploring the microstructure of children's central conceptual structures in the domain of number. *Monographs of the Society for Research in Child Development, 61*(1-2, Serial No. 246), 27-58. https://doi.org/10.2307/1166077

Orlov, T., Yakovlev, V., Hochstein, S., & Zohary, E. (2000). Macaque monkeys categorize images by their ordinal number. *Nature, 404*(6773), 77–80. https://doi.org/10.1038/35003571

Palmer, A., & Baroody, A. J. (2011). Blake's development of the number words "one," "two," and "three." *Cognition and Instruction, 29*(3), 265–296. https://doi.org/10.1080/07370008.2011.583370

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., ... & Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ, 372*(71). https://doi.org/10.1136/bmj.n71

Paras, J. (2001). Crisis in mathematics education. Student failure: challenges and possibilities. *South African Journal of Higher Education, 15*(3), 66-73. https://doi.org/10.4314/sajhe.v15i3.25327

Pastore, N. (1961). Number sense and 'counting' ability in the canary. *Zeitschrift für Tierpsychologie, 18*(5), 561–573. https://doi.org/10.1111/j.1439-0310.1961.tb00242.x

Peters, M. D. J., Godfrey, C. M., McInerney, P., Munn, Z., Tricco, A. C., & Khalil, H. (2020). Chapter 11: Scoping Reviews (2020 version). In E. Aromataris & Z. Munn (Eds.), *JBI manual for evidence synthesis*. JBI. https://doi.org/10.46658/JBIMES-20-12

Peters, M. D. J., Godfrey, C., McInerney, P., Munn, Z., Tricco, A. C., & Khalil, H. (2022). Best practice guidance and reporting items for the development of scoping review protocols. *JBI Evidence Synthesis, 20*(4), 953–968. https://doi.org/10.11124/JBIES-21-00242

Piaget, J. (1954). *The construction of reality in the child* (M. Cook, Trans.). Basic Books. https://doi.org/10.1037/11168-000

Piaget, J. (1970). *Science of education and the psychology of the child*. Orion.

Press. https://doi.org/10.2307/1423801

Piaget, J. (1971). *Biology and knowledge: An essay on the relations between organic regulations and cognitive processes*. Chicago, IL: University of Chicago Press. https://doi.org/10.1086/407227

Piaget, J., & Szeminska, A. (1941). The child's conception of number. Selected pages reprinted in Gruber, H. E., & Voneche, J. J. (1977). *The essential Piaget* (pp. 298–311). https://doi.org/10.4324/9781315009643

Piazza, M., Giacomini, E., Le Bihan, D., & Dehaene, S. (2003). Single-trial classification of parallel pre-attentive and serial attentive processes using functional magnetic resonance imaging. *Proceedings of the Royal Society of London. Series B: Biological Sciences, 270*(1521), 1237–1245. https://doi.org/10.1098/rspb.2003.2356

Piazza, M., Mechelli, A., Butterworth, B., & Price, C. J. (2002). Are subitising and counting implemented as separate or functionally overlapping processes? *NeuroImage, 15*(2), 435–446. https://doi.org/10.1006/nimg.2001.0980

Pimm, D., & Keynes, M. (1994). Mathematics classroom language: Form, function and force. In R. Biehler, R. W. Scholz, R. Sträßer, & B. Winkelmann (Eds.), *Didactics of mathematics as a scientific discipline* (pp. 159–169). Kluwer Academic Publishers.

Power, R. J., & Dal Martello, M. F. (1990). The dictation of Italian numerals. *Language and Cognitive Processes, 5*(3), 237–254. https://doi.org/10.1080/01690969008402106

Prast, E. J., Van de Weijer-Bergsma, E., Kroesbergen, E. H., & Van Luit, J. E. (2018). Differentiated instruction in primary mathematics: Effects of teacher professional development on student achievement. *Learning and Instruction, 54*, 22-34. https://doi.org/10.1016/j.learninstruc.2018.01.009

Rameli, M. R. M., & Kosnin, A. M. (2016). *Challenges in mathematics learning: a*

study from school students' perspective. Universiti Teknologi Malaysia.

Renkl, A. (2014). Toward an instructionally oriented theory of example-based learning. *Cognitive Science, 38*(1), 1-37. https://doi.org/10.1111/cogs.12086

Reynvoet, B., & Brysbaert, M. (1999). Single-digit and two-digit Arabic numerals address the same semantic number line. *Cognition, 72*(2), 191–201. https://doi.org/10.1016/s0010-0277(99)00048-7

Rischel, J. (2012). Danish. *Revue belge de Philologie et d'Histoire, 90*(3), 809–832.

Roberts, G. (2000). Bilingualism and number in Wales. *International Journal of Bilingual Education and Bilingualism, 3*(1), 44–56. https://doi.org/10.1080/13670050008667699

Roberts, S. A., & Olarte, T. R. (2025). Enacting multilingual learner core practices: A PST's approximations of practice of mathematics language routines. *Journal of Mathematics Teacher Education, 28*(3), 465-490. https://doi.org/10.1007/s10857-023-09600-8

Roberts, W. A., & Mitchell, S. (1994). Can a pigeon simultaneously process temporal and numerical information? *Journal of Experimental Psychology: Animal Behavior Processes, 20*(1), 66. https://doi.org/10.1037//0097-7403.20.1.66

Rodríguez-Martínez, J. A., González-Calero, J. A., del Olmo-Muñoz, J., Arnau, D., & Tirado-Olivares, S. (2023). Building personalised homework from a learning analytics based formative assessment: Effect on fifth-grade students' understanding of fractions. *British Journal of Educational Technology, 54*(1), 76-97 https://doi.org/10.1111/bjet.13292

Rogers, C. R. (1957*). Becoming a person*. Association Press

Rothstein, H. R., Sutton, A. J., & Borenstein, M. (Eds.). (2005). *Publication bias in meta-analysis: Prevention, assessment and adjustments.* John Wiley & Sons. https://doi.org/10.1002/0470870168

Rumbaugh, D. M., Savage-Rumbaugh, S., & Hegel, M. T. (1987). Summation in the

chimpanzee (Pan troglodytes). Journal of Experimental Psychology: *Animal Behavior Processes, 13*(2), 107. https://doi.org/10.1037//0097-7403.13.2.107

Russo, J. (2017). Sensible and crazy numbers. *Australian Primary Mathematics Classroom, 22*(2), 17–20

Sarnecka, B. W., & Gelman, S. A. (2004). Six does not just mean a lot: Preschoolers see number words as specific. *Cognition, 92*(3), 329–352. https://doi.org/10.1016/j.cognition.2003.10.001

Sarnecka, B. W., & Lee, M. D. (2009). Levels of number knowledge during early childhood. *Journal of Experimental Child Psychology, 103*(3), 325–337. https://doi.org/10.1016/j.jecp.2009.01.002

Sarnecka, B. W., Goldman, M. C., & Slusser, E. B. (2015). How counting leads to children's first representations of exact, large numbers. In R. Cohen Kadosh & A. Dowker (Eds.), *The Oxford handbook of numerical cognition* (pp. 291–309). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199642342.013.011

Sarnecka, B. W., Kamenskaya, V. G., Yamana, Y., Ogura, T., & Yudovina, Y. B. (2007). From grammatical number to exact numbers: Early meanings of "one," "two," and "three" in English, Russian, and Japanese. *Cognitive Psychology, 55*(2), 136–168. https://doi.org/10.1016/j.cogpsych.2006.09.001

Sathian, K., Simon, T. J., Peterson, S., Patel, G. A., Hoffman, J. M., & Grafton, S. T. (1999). Neural evidence linking visual object enumeration and attention. *Journal of Cognitive Neuroscience, 11*(1), 36–51. https://doi.org/10.1162/089892999563212

Saxe, G. B. (1981). Body parts as numerals: A developmental analysis of numeration among the Oksapmin in Papua New Guinea. *Child Development, 52*(2), 306–316. https://doi.org/10.2307/1129249

Saxton, M., & Towse, J. N. (1998). Linguistic relativity: The case of place value in

multi-digit numbers. *Journal of Experimental Child Psychology, 69*(1), 66–79. https://doi.org/10.1006/jecp.1998.2434

Schaeffer, B., Eggleston, V. H., & Scott, J. L. (1974). Number development in young children. *Cognitive Psychology, 6*(3), 357–379. https://doi.org/10.1016/0010-0285(74)90017-6

Scholl, B. J., & Leslie, A. M. (1999). Explaining the infant's object concept: Beyond the perception/cognition dichotomy. In M. S. Gazzaniga (Ed.), *What is cognitive science?* (pp. 26–73). MIT Press.

Setati, M., Chitera, N., & Essien, A. (2009). Research on multilingualism in mathematics education in South Africa: 2000–2007. *African Journal of Research in Mathematics, Science and Technology Education, 13*(sup1), 65-80. https://doi.org/10.1080/10288457.2009.10740662

Sfard, A. (2001). There is more to discourse than meets the ears: Looking at thinking as communicating to learn more about mathematical learning. *Educational Studies in Mathematics*, 46(1), 13-57. https://doi.org/10.1007/0-306-48085-9_1

Shaheen, N., Shaheen, A., Ramadan, A., Hefnawy, M. T., Ramadan, A., Ibrahim, I. A., Hassanein, M. E., Ashour, M. E., & Flouty, O. (2023). Appraising systematic reviews: A comprehensive guide to ensuring validity and reliability. *Frontiers in Research Metrics and Analytics, 8*, 1268045. https://doi.org/10.3389/frma.2023.1268045

Shaki, S., & Fischer, M. H. (2008). Reading space into numbers: A cross-linguistic comparison of the SNARC effect. *Cognition, 108*(2), 590–599. https://doi.org/10.1016/j.cognition.2008.05.001

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

Siegler, R. S., & Mu, Y. (2008). Chinese children excel on novel mathematics

problems even before elementary school. *Psychological Science, 19*(8), 759–763. https://doi.org/10.1111/j.1467-9280.2008.02153.x

Siegler, R. S., & Opfer, J. E. (2003). The development of numerical estimation: Evidence for multiple representations of numerical quantity. *Psychological Science, 14*(3), 237–250. https://doi.org/10.1111/1467-9280.02438

Siegler, R. S., Thompson, C. A., & Opfer, J. E. (2009). The logarithmic-to-linear shift: One learning sequence, many tasks, many time scales. *Mind, Brain, and Education, 3*(3), 143–150. https://doi.org/10.1111/j.1751-228X.2009.01064.x

Simon, T. J. (1997). Reconceptualizing the origins of number knowledge: A "non-numerical" account. *Cognitive Development, 12*(3), 349–372.

Simon, T. J., Hespos, S. J., & Rochat, P. (1995). Do infants understand simple arithmetic? A replication of Wynn (1992). *Cognitive Development, 10*(2), 253–269. https://doi.org/10.1016/0885-2014(95)90003-9

Skwarchuk, S. L., Sowinski, C., & LeFevre, J. A. (2014). Formal and informal home learning activities in relation to children's early numeracy and literacy skills: The development of a home numeracy model. *Journal of Experimental Child Psychology, 121*, 63–84. https://doi.org/10.1016/j.jecp.2013.11.006

Slaughter, V., Itakura, S., Kutsuki, A., & Siegal, M. (2011). Learning to count begins in infancy: Evidence from 18-month-olds' visual preferences. *Proceedings of the Royal Society B: Biological Sciences, 278*(1720), 2979–2984. https://doi.org/10.1098/rspb.2011.0227

Smale-Jacobse, A. E., Meijer, A., Helms-Lorenz, M., & Maulana, R. (2019). Differentiated instruction in secondary education: A systematic review of research evidence. *Frontiers in Psychology, 10*, 2366. https://doi.org/10.3389/fpsyg.2019.02366

Smirnova, A. A., Lazareva, O. F., & Zorina, Z. A. (2000). Use of number by crows: Investigation by matching and oddity learning. *Journal of the Experimental*

Analysis of Behavior, 73(2), 163–176. https://doi.org/10.1901/jeab.2000.73-163

Smit, J. (2013). *Scaffolding language in multilingual mathematics classrooms* [Doctoral dissertation, Utrecht University]. Utrecht University Repository. https://dspace.library.uu.nl/handle/1874/275867

Snyder, H. (2019). Literature review as a research methodology: An overview and guidelines. *Journal of Business Research, 104*, 333–339. https://doi.org/10.1016/j.jbusres.2019.07.039

Song, M. J., & Ginsburg, H. P. (1988). The effect of the Korean number system on young children's counting: A natural experiment in numerical bilingualism. *International Journal of Psychology, 23*(1-6), 319–332. https://doi.org/10.1080/00207598808247762

Sousa, D. A. (2008). *Facilitator's guide: How the brain learns mathematics*. Corwin Press.

Sowder, J. T., & Wheeler, M. M. (1989). The development of concepts and strategies used in computational estimation. *Journal for Research in Mathematics Education, 20*(2), 130–146. https://doi.org/10.5951/jresematheduc.20.2.0130

Starkey, P., & Cooper, R. G. (1980). Perception of numbers by human infants. *Science, 210*(4473), 1033–1035. https://doi.org/10.1126/science.7434014

Starkey, P., Spelke, E. S., & Gelman, R. (1990). Numerical abstraction by human infants. *Cognition, 36*(2), 97–127. https://doi.org/10.1016/0010-0277(90)90001-Z

Strauss, M. S., & Curtis, L. E. (1981). Infant perception of numerosity. *Child Development, 52*(4), 1146–1152. https://doi.org/10.2307/1129497

Sun, X. (2015). Chinese core tradition to whole number arithmetic. In X. Sun, B. Kaur, & J. Novotná (Eds.), *Proceedings of the Twenty-Third ICMI Study: Primary Mathematics Study on Whole Numbers* (pp. 22–28). Macau, China: University of Macau

Sun, X. H., & Bussi, M. G. B. (2018). Language and cultural issues in the teaching and learning of WNA. In M. G. Bartolini Bussi & X. Sun (Eds.), *Building the foundation: whole numbers in the primary grades* (pp. 35–70). Springer. https://doi.org/10.1007/978-3-319-63555-2_3

Suprayogi, M. N., Valcke, M., & Godwin, R. (2017). Teachers and their implementation of differentiated instruction in the classroom. *Teaching and Teacher Education, 67*, 291-301. https://doi.org/10.1016/j.tate.2017.06.020

Sutton, A., Clowes, M., Preston, L., & Booth, A. (2019). Meeting the review family: Exploring review types and associated information retrieval requirements. *Health Information & Libraries Journal, 36*(3), 202–222. https://doi.org/10.1111/hir.12276

Suurtamm, C., Koch, M. J., & Arden, A. (2010). Teachers' assessment practices in mathematics: Classrooms in the context of reform. *Assessment in Education: Principles, Policy & Practice, 17*(4), 399–417. https://doi.org/10.1080/0969594X.2010.496546

Suzuki, K., & Kobayashi, T. (2000). Numerical competence in rats (Rattus norvegicus). *Animal Cognition, 3*(2), 113–121.

Sweller, J., Ayres, P., & Kalyuga, S. (2011). Measuring cognitive load. In *Cognitive load theory* (pp. 71–85). Springer. https://doi.org/10.1007/978-1-4419-8126-4_6

Terrace, H. S. (2014). Animal cognition. In H. S. Terrace (Ed.), *Animal cognition* (pp. 7–28). Psychology Press. https://doi.org/10.4324/9781315802602

Thanheiser, E. (2009). Preservice elementary school teachers' conceptions of multidigit whole numbers. *Journal for Research in Mathematics Education, 40*(3), 251-281. https://doi.org/10.5951/jresematheduc.40.3.0251

Thanheiser, E., & Melhuish, K. (2019). Leveraging variation of historical number systems to build understanding of the base-ten place-value system. *ZDM, 51*(1),

39-55. https://doi.org/10.1007/s11858-018-0984-7

Tomlinson, C. A. (2001). *How to differentiate instruction in mixed-ability classrooms* (2nd ed.). ASCD.

Tomlinson, C. A. (2014). *The differentiated classroom: Responding to the needs of all learners* (2nd ed.). ASCD. https://doi.org/10.63467/red.13.2025.art1

Tricco, A. C., Lillie, E., Zarin, W., O'Brien, K. K., Colquhoun, H., Levac, D., & Straus, S. E. (2018). PRISMA extension for scoping reviews (PRISMA-ScR): Checklist and explanation. *Annals of Internal Medicine, 169*(7), 467–473. https://doi.org/10.7326/M18-0850

Tsertsvadze, A., Chen, Y.-F., Moher, D., Sutcliffe, P., & McCarthy, N. (2015). How to conduct systematic reviews more expeditiously? *Systematic Reviews, 4*(1), 1–12. https://doi.org/10.1186/s13643-015-0147-7

Turkan, S., & Liu, O. L. (2012). Differential performance by English language learners on an inquiry-based science assessment. *International Journal of Science Education, 34*(15), 2343–2369. https://doi.org/10.1080/09500693.2012.697208

Turkan, S., de Oliveira, L. C., Lee, O., & Phelps, G. (2014). Proposing a knowledge base for teaching academic content to English language learners: Disciplinary linguistic knowledge. *Teachers College Record, 116*(3), 1–30. https://doi.org/10.1177/016146811411600303

Udry, I., & Berthele, R. (2019). *Dreisprachige Primarschulen: Theoretische Grundlagen und Umsetzungsmöglichkeiten.* Bericht des Instituts für Mehrsprachigkeit, 5–9.

Uller, C., Carey, S., Huntley-Fenner, G., & Klatt, L. (1999). What representations might underlie infant numerical knowledge? *Cognitive Development, 14*(1), 1–36. https://doi.org/10.1016/S0885-2014(99)80016-1

UNESCO (2016). *Global Education Monitoring Report Policy Paper 24: If you don't*

understand, how can you learn? Paris: UNESCO.

Uttal, D. H., Scudder, K. V., & DeLoache, J. S. (1997). Manipulatives as symbols: A new perspective on the use of concrete objects to teach mathematics. *Journal of Applied Developmental Psychology, 18*(1), 37-54. https://doi.org/10.1016/s0193-3973(97)90013-7

Van Loosbroek, E., & Smitsman, A. W. (1990). Visual perception of numerosity in infancy. *Developmental Psychology, 26*(6), 916–922. https://doi.org/10.1037/0012-1649.26.6.916

Van Oers, B. (2001). Contextualisation for abstraction. *Cognitive Science Quarterly, 1*(3), 279–305.

Verguts, T., & Fias, W. (2006). Lexical and syntactic structures in a connectionist model of reading multi-digit numbers. *Connection Science, 18*(3), 265–283. https://doi.org/10.1080/09540090600821976

Verra, R. (2016). Ladin: *The Ladin language in education in Italy*. Regional Dossiers Series. Mercator European Research Centre on Multilingualism and Language Learning, Fryske Akademy.

Villarroel, J. D., Angós, T., & Grandmontagne, A. (2009). The origin of numerical concepts: Early meanings of "one," "two," and "three" among Basque- and Spanish-speaking children. *Problems of Education in the 21st Century, 10*, 109–118. https://doi.org/10.1007/s10649-010-9291-0

von Mengden, F. (2005). How myths persist: Jacob Grimm, the long hundred, and duodecimal counting. In G. Knappe (Ed.), *Englische Sprachwissenschaft und Mediävistik: Standpunkte - Perspektiven - Neue Wege* (pp. 201–222). Frankfurt am Main: Peter Lang.

Vygotsky, L. S. (1962). *Thought and language* (E. Hanfmann & G. Vakar, Trans.). MIT Press. (Original work published 1934) https://doi.org/10.1037/11193-000

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological*

processes. Harvard University Press. https://doi.org/10.1525/aa.1979.81.4.02a00580

Wagner, A. R. (2014). SOP: A model of automatic memory processing in animal behavior. In A.R. Wagner (Ed.), *Information processing in animals* (pp. 5–47). Psychology Press. https://doi.org/10.1007/978-3-319-47829-6_1545-1

Wassmann, J., & Dasen, P. R. (1994). Yupno number system and counting. *Journal of Cross-Cultural Psychology, 25*(1), 78–94. https://doi.org/10.1177/0022022194251005

Wege, T. E., Bourque, T., Merkley, R., & Cheung, P. (2023). Thirty years of knower-levels: A systematic review of the Give-N task. *PsyArXiv.* https://doi.org/10.31234/osf.io/fh95s

Werner, H. (1957). The concept of development from a comparative and organismic point of view. In D. B. Harris (Ed.), *The concept of development* (pp. 125–148). Minneapolis, MN: University of Minnesota Press.

West, R. E., & Young, R. J. (2002). Do domestic dogs show any evidence of being able to count? *Animal Cognition, 5*(3), 183–186. https://doi.org/10.1007/s10071-002-0137-7

Whalen, J., Gallistel, C. R., & Gelman, R. (1999). Nonverbal counting in humans: The psychophysics of number representation. *Psychological Science, 10*(2), 130–137. https://doi.org/10.1111/1467-9280.00120

Widom, T. R., & Schlimm, D. (2012). Methodological reflections on typologies for numerical notations. *Science in Context, 25*(2), 155–195. https://doi.org/10.1017/S0269889712000031

Wiese, H. (2003). *Numbers, language, and the human mind*. Cambridge University Press. https://doi.org/10.1017/cbo9780511486562

Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry, 17*(2), 89–100.

https://doi.org/10.1111/j.1469-7610.1976.tb00381.x

Wynn, K. (1990). Children's understanding of counting. *Cognition, 36*(2), 155–193. https://doi.org/10.1016/0010-0277(90)90003-3

Wynn, K. (1992a). Addition and subtraction by human infants. *Nature, 358*(6389), 749–750. https://doi.org/10.1038/358749a0

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

Wynn, K. (1996). Infants' individuation and enumeration of actions. *Psychological Science, 7*(3), 164–169. https://doi.org/10.1111/j.1467-9280.1996.tb00348.x

Wynn, K. (1998). Psychological foundations of number: Numerical competence in human infants. *Trends in Cognitive Sciences, 2*(8), 296–303. https://doi.org/10.1016/S1364-6613(98)01203-0

Wynn, K., Bloom, P., & Chiang, W. C. (2002). Enumeration of collective entities by 5-month-old infants. *Cognition, 83*(3), B55–B62. https://doi.org/10.1016/S0010-0277(02)00008-2

Yi, J., Son, J. W., & Chang, H. (2025). Teacher positioning impact on students' learning opportunities in multilingual mathematics classrooms. *Journal of Mathematics Teacher Education*, 1-30. https://doi.org/10.1007/s10857-025-09704-3

Zebian, S. (2005). Linkages between number concepts, spatial thinking, and directionality of writing: The SNARC effect and the reverse SNARC effect in English and Arabic monoliterates, biliterates, and illiterate Arabic speakers. *Journal of Cognition and Culture, 5*(1-2), 165–190. https://doi.org/10.1163/1568537054068660

Zerai, D., Eskelä-Haapanen, S., Posti-Ahokas, H., & Vehkakoski, T. (2023). The use of question modification strategies to differentiate instruction in Eritrean

mathematics and science classrooms. *Education Sciences, 13*(3), 284. https://doi.org/10.3390/educsci13030284

Zhang, D., Ding, Y., Lee, S., & Chen, J. (2017). Strategic development of multiplication problem solving: Patterns of students' strategy choices. *Journal of Educational Research, 110*(2), 159-170. https://doi.org/10.1080/00220671.2015.1060928

Zhang, J., & Norman, D. A. (1995). A representational analysis of numeration systems. *Cognition, 57*(3), 271–295. https://doi.org/10.1016/0010-0277(95)00668-2

Zhang, L., Lei, Y., Pelton, T., Pelton, L. F., & Shang, J. (2024). An exploration of gendered differences in cognitive, motivational and emotional aspects of game-based math learning. *Journal of Computer Assisted Learning, 40*(6), 2633-2649. https://doi.org/10.1111/jcal.12956

Ziegler, E., Edelsbrunner, P. A., & Stern, E. (2021). The benefit of combining teacher-direction with contrasted presentation of algebra principles. *European Journal of Psychology of Education, 36*(1), 187-218. https://doi.org/10.1007/s10212-020-00468-3

Zuber, J., Pixner, S., Moeller, K., & Nuerk, H. C. (2009). On the language specificity of basic number processing: Transcoding in a language with inversion and its relation to working memory capacity. *Journal of Experimental Child Psychology, 102*(1), 60–77. https://doi.org/10.1016/j.jecp.2008.04.003

Appendix

Appendix 1: Letter confirming Research Ethics Approval

Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

Application	Academic Year 2023/24
Applicant Code	MT18
Applicant/Supervisor Name	Insa-Annett Tiaden / Prof. Lorna Carson
Title of Research	The Influence of Language Properties on the Acquisition of Numerical Skills in Children
Date of this letter	02.11.2023

Dear Insa-Annett,

Your submission for ethical approval for the research project above was considered by the Research Ethics Committee, School of Linguistic, Speech and Communication Sciences, Trinity College Dublin on 02/11/2023 and has been approved in full.

Please note:
(i) On completion of research projects, applicants should complete the *End of Project Report Form* (which can be found at: https://www.tcd.ie/slscs/research/ethics/) and submit one electronic copy (to slscs@tcd.ie)
(ii) The REC requests, in particular, that you attend to your commitments regarding the storage and destruction of data arising from this research, in keeping with REC policy and General Data Protection Regulation (GDPR) guidelines.

We wish you every luck with your research.

Best wishes,

Bronagh Ćatibušić

Dr Bronagh Ćatibušić
Chair, Research Ethics Committee
School of Linguistic, Speech and Communication Sciences

Scoil na nEolaíochtaí Teangeolaíochta,
Urlabhra agus Cumarsáide,
Coláiste na Tríonóide,
Baile Átha Cliath 2, Éire

School of Linguistic, Speech &
Communication Sciences,
Trinity College,
Dublin 2, Ireland

T 353 (0)1 896 1560
slscs@tcd.ie
www.tcd.ie/slscs

Appendix 2: Scoping Review Summary Table

Details	Study #1	Study #2	Study #3	Study #4
Author (Year)	**Cankaya et. al (2014)**	**Clayton et. al (2020)**	**Dowker & Roberts (2015)**	**Helmreich et. al (2011)**
Citation	Cankaya, O., LeFevre, J.-A., & Sowinski, C. (2014). The role of number naming systems and numeracy experiences in children's rote counting: Evidence from Turkish and Canadian children. Learning and Individual Differences, 32, 238–245.	Clayton, F., Copper, C., Steiner, A. F., Banfi, C., Finke, S., Landerl, K., & Göbel, S. (2020). Two-digit number writing and arithmetic in Year 1 children: Does number word inversion matter? Cognitive Development, 56, 100967.	Dowker, A., & Roberts, R. (2015). Does the transparency of the counting system affect children's numerical abilities? Frontiers in Psychology, 6, 945.	Helmreich, I., Zuber, J., Pixner, S., Kaufmann, L., Nuerk, H.-C., & Moeller, K. (2011). Language effects on children's nonverbal number line estimations. Journal of Cross-Cultural Psychology, 42(4), 598–613.
Abstract	How important is the regularity of a number naming system when young children are learning to count? Three-and four-year-old English- and Turkish-speaking children (N = 75) played one of two versions of a board game. The number game involved counting and naming numbers from 1 to 20 whereas the color game involved naming colors. All of the Turkish-speaking children in the number game condition showed some improvement in their rote counting skills whereas almost half of the English-speaking children did not. Despite the advantage of the regular number naming system, however, the Turkish-speaking children had lower scores than the English-speaking children on other early numeracy tasks. These results support the view that both the characteristics of the number language and factors such as children's numeracy-related experiences must be considered when comparing children's early numeracy skills across language groups.	Children's understanding of place-value structure is essential for later mathematics. Some languages (e.g., English) have an inversion property in two-digit names (e.g., "twenty-one" vs. "one-and-twenty"), which could influence children's number writing and arithmetic. In this study, Year 1 children speaking languages with and without inversion (English vs. German/Italian) were assessed on two-digit number writing and arithmetic tasks. Results showed that children from the non-inversion language groups outperformed those from the inversion language group on writing two-digit numerals, but inversion did not significantly affect arithmetic performance once numeral knowledge was accounted for. These findings suggest that number word inversion may influence early numeral writing, but does not directly constrain broader arithmetic learning in young children.	The transparency of number naming systems differs across languages, and this may influence children's numerical development. In this study, we compared children speaking languages with transparent counting systems (e.g., Turkish) and less transparent systems (e.g., English) on a range of numerical tasks, including counting, number comparison, and arithmetic. Results showed that transparency of the counting system was associated with some advantages in basic number processing tasks, although these effects varied depending on the specific skill measured. This suggests that linguistic structure may play a role in early numerical abilities, but language transparency does not fully account for all differences in number cognition.	Language influences the way children represent numerical magnitude on nonverbal number line tasks. This study investigated how linguistic factors affect children's estimations of numerical magnitudes on a number line. German- and Czech-speaking 5- to 8-year-olds estimated positions of numbers on a 0–100 number line. Results showed that language characteristics, particularly the transparency of number words and the directionality of writing systems, influenced children's estimation patterns. Specifically, children speaking languages with more transparent number naming exhibited more linear estimation patterns. These findings suggest that language properties affect nonverbal numerical representations in early childhood.
Methodology	Number game inv. counting and naming numbers	Test of number writing and concurrent arithmetic Performance	Nonverbal MNL estimation and transcoding	Nonverbal MNL estimation and general cognitive abilities test
Purpose	To examine the influence of different number naming systems and numeracy experiences on young children's acquisition of early numeracy skills, particularly rote counting.	1: Shed more light onto the influence of language on transcoding of two-digit numbers 2: evaluate the relationship between number writing and arithmetic in more detail. In particular, we aimed to compare the relative contribution of two-digit number writing as a concurrent predictor of arithmetic across the two languages, German with inverted number words and English with non-inverted number words.	Investigate if language properties influence children's performance in certain numerical tasks	To investigate how language-specific number word properties, particularly inversion in German and partial inversion in Italian, influence first graders' nonverbal number line estimation and mental representation of numerical magnitude.

Research Questions	How important is the regularity of a number naming system when young children are learning to count?	Do differences in NWS influence transcoding between spoken word and Arabic digits and its relationship with maths?	Does the language of school instruction affect the MNL in children who have other wise similar cultural and educational experiences?	Does language, as reflected by the structure of its number word system, influences performance in the non-verbal spatial-numerical number line estimation task?

Details	Study #1	Study #2	Study #3	Study #4
Hypotheses	H1: "We hypothesized that when children who did not have much prior counting knowledge were exposed systematically to number names and the counting string, a relatively regular number naming system (Turkish) would facilitate children's acquisition of counting words more than an irregular number naming system (English). H2: " Also, we expected that number game exposure, relative to colour game exposure, would facilitate early numeracy skills for both English- and Turkish-speaking children, because previous research has shown benefits on various numerical skills including counting"	The inverted structure of teen number words in German and English will lead to inversion errors in both language groups, German-speaking children will show more inversion errors for two-digit numbers larger than 20 and fewer non-inversion errors for two-digit numbers larger than 20, than English-speaking children. Furthermore, number writing of two-digit numbers would be a significant predictor of concurrent performance in arithmetic in both languages.	H1: The Welsh-medium children would score higher on the number line estimation tasks, and this would especially be true of the number line from 1 to 100, as this places greater demands on the ability to represent multi-digit numbers, for which a transparent counting system would provide an advantage. H2: The Welsh-medium children would perform better on the tasks involving reading and writing numbers, but there would be no difference between the two groups in BAS Number Skills test performance, based on Dowker et al.'s (2008) earlier findings and their suggestion that linguistic effects on mathematics may be specific rather than global.	H1: It was expected that the reversed order of tens and units in German increases the estimation error in Austrian children. H2: Apart from the log-to-linear assumption, children's overall estimation performance may also be explained by a two-linear model suggesting that developing a mental representation of number magnitude incorporates the mastery of the place-value structure of the Arabic number system. H3: In the context of place-value integration, German-speaking children's estimations should be particularly erroneous for numbers for which mixing up tens and units results in a large estimation error (e.g., 82 → 28) compared to numbers leading to only a small estimation error (e.g., 54 → 45). H3a: Related to this, an even more specific hypothesis can be made: German-speaking children should specifically underestimate numbers such as 82 where mixing up tens and units results in a smaller number (→ 28). Contrarily, they should overestimate numbers such as 27 for which mixing up tens and units results in larger number (→ 72). As there is no inversion in Italian number words, this under-/overestimation should be more pronounced in German-speaking children.
Languages	English, Turkish	German , English	Welsh, English	German, Italian
Assigned condition	English: colour game: 8 boys, 8 girls; number board game : 6 b., 8 g. Turkish: colour game: 12 b., 9 g.; number board game : 10 b., 14 g.	language groups by native country/language	language groups by native language/language of instruction	language groups by native country/language
Mono-/Bilingual	All Turkish speaking children were monolingual, the Canadian children spoke English at day-care, however 43% also spoke another language at home.	not specified	While all lessons at the Welsh-medium school were conducted exclusively in Welsh, English was the first language of 13 of the 20 children attending that school.	not specified
Setting	Kindergarten and daycare	Primary school	Classroom	Primary school
Country	Ottawa, Canada and Istanbul, Turkey	Yorkshire, UK and Graz, Austria	Cardiff and Rhondda Cynon Taf area, Wales	Austria, Italy

Number Naming Variations	Regular (Turkish) vs less regular (English) NNS	Fully-inverted (German) vs minimally-inverted English)	Regular (Welsh) vs less regular (English) NNS	Fully-inverted (German) vs minimally-inverted (Italian)

Details	Study #1	Study #2	Study #3	Study #4
Setup/ design	The setup contained a bridge made up of squares in the repeated colour sequence red, yellow, green, and blue. Next to the board were spinners divided into four squares in yellow, red, green, and blue, respectively. Both game setups were identical except that the number game had the numbers 1 – 20 printed inside the squares of the bridge and the numbers 1 and 2 twice each inside the spinner squares. After spinning the spinner the child was allowed to move their token to the colour or indicated or either 1 or 2 squares ahead depending on the number the spinner landed on. The child was asked to say the colours or numbers inside the squares the token moved they past out loud	The participants were assessed on four tasks, number writing, arithmetic and non-verbal reasoning.	The children completed four tasks: the BAS Number Skills test, which is a standardized test that assesses written calculation (Elliott et al., 1997), two transcoding tasks (writing to dictation and reading aloud), and a non-verbal number line estimation task.	The experiment reported on in the given article was part of a larger cross-cultural project investigating the early development of numeracy in children with different number words systems. Children in Italy and Austria were examined in a variety of numerical and non-numerical tasks. However, the article reviewed here focuses on the nonverbal number line estimation task only.
Pre-/post-test	Children met individually with the experimenter. The pre-test sessions consisted of five tasks: Rote counting, Non-symbolic arithmetic task, KeyMath numeration subtest (Conolly, 2000) (standardized test includes a variety of early, numeracy tasks, including counting, ordinality,	not applicable		not applicable
Testing procedure	In total there were four game sessions for which every child met with an experimenter once a week over four weeks. During these four sessions each child played the games approximately 8 – 10 times. No game lasted more than 15 minutes with the average being 3 – 8 minutes. Before each session children were asked to rote count as high as they could. Before the third game session children additionally completed the number line estimation tasks and the number recognition task. No feedback was given at any point by the experimenter	Four assessed tasks were part of a larger battery of measures administered in a fixed order in a classroom setting in two 1-h sessions. Working memory was tested individually. All tests were administered in English for the English-speaking and in German for the German-speaking children.	Children were tested individually in single sessions, with instructions in Welsh for Welsh-medium and English for English-medium education children. Trials were presented sequentially without performance feedback, including practice trials. In the writing-to-dictation task, participants wrote 32 verbally presented Arabic numbers (2 single-digit, 10 double-digit, 20 three-digit). In the reading-aloud task, participants read aloud 32 Arabic numerals (same distribution) displayed one by one on a computer screen. Both transcoding tasks scored responses as 1 for correct and 0 for incorrect. The number line estimation task required children to estimate visually presented numbers on empty 10 cm paper number lines labeled "0" (left) and "20" or "100" (right), without counting or other strategies. Numbers estimated on 0–20 lines were 12, 1, 13, 4, 15, 19, 7, 17, and 5; on 0–100 lines were 27, 2, 64, 35, 7, 13, 99, 75, 47, 3, 11, 82, 95, 9, 17, 6, 18, and 53. Before starting, participants completed practice estimations of 10 on the 0–20 line and 50 on the 0–100 line.	Children completed a paper-pencil number-to-position task individually, estimating the location of numbers (printed in Arabic notation) on a 0–100 empty number line. A 50 trial practice was given, with instructions to estimate only, without counting or feedback, to examine the effect of language properties (e.g., German number-word inversion) on nonverbal number line estimation.
Scoring	All in all, each child had six attempts to count as high as possible, once during pre-test, four times during the game sessions and once during post-test. The resulting data were analysed to see if patters of change were emerging due to training. Scores were given to the children according to highest count achieved and in which session it occurred. The researchers deemed this a better reflection of progress than just taking pre- and post-test performance into account, since counting performance was highly variable across all sessions including post-test. Then the participants were grouped according to	Number writing task (52 dictated items; teens and two-digit numbers >20, excluding decades); two-digit responses scored correct/incorrect and total accuracy calculated. Arithmetic assessed using WIAT-II UK Numerical Operations (6 digit/dot items + 9 increasing-difficulty calculations; 15 min). Non-verbal reasoning measured with Raven's SPM Plus (Sets A–C; 34 scored items; 15 min; max = 34). Working memory assessed individually via Forward/Backward Digit Recall and Forward/Backward Block Recall (WMTB-C), each beginning with practice trials to determine span.	To obtain number line estimation scores, the distance between the true position of the number that was presented and the position of the number corresponding to the child's estimate on the number line was measured to the nearest millimetre. These deviation measures were then averaged for each participant individually to give two mean estimation error scores: one score for the 0–20 number line estimations, and one for the 0–100 number line estimations. These were also averaged to obtain an overall mean estimation error for each participant.	The deviation of estimated position from the true position for every single item was measured automatically to the nearest millimetre and the individual deviation of the estimated from the actual value was computed for each child separately to obtain estimation accuracy.

Details	Study #1	Study #2	Study #3	Study #4
Key findings	Here is a concise summary suitable for a scoping review summary table cell: Intervention Effects: Turkish-speaking children benefitted more from the training than English-speaking children, with most reaching their highest counts post-intervention (11% pre, 36% during, 53% after), whereas English-speaking children most often peaked pre-intervention (47% pre). Improvement depended on condition: children in the number game condition showed substantial gains (66% post; all Turkish and half of English improved), while the colour game condition showed little improvement (24% post). Across other measures, English-speaking children outperformed in cardinality, KeyMath, and number recognition, while Turkish-speaking children performed better in non-symbolic arithmetic. Overall	Controlling for age (German-speaking children significantly older), no significant group differences were found in arithmetic, non-verbal reasoning, or working memory. Overall, inversion errors comprised 81% of errors in the German group (430 vs. 98 non-inversion) but only 21% in the English group (76 vs. 292 non-inversion). For teens, inversion error proportions did not differ significantly between groups. For two-digit numbers, German-speaking children showed significantly more inversion than non-inversion errors, whereas English-speaking children showed the opposite pattern. Germans made more inversion errors for two-digit numbers than teens; English speakers showed more inversion for teens than two-digit numbers.	BAS & Number Line Estimation (MNL): Age did not significantly correlate with BAS scores or estimation errors (trend toward better 0–100 performance with age), but correlated strongly with Reading Aloud and Writing. BAS raw and standard scores negatively correlated with overall and 0–100 estimation errors (not 0–20) and positively with Reading Aloud and Writing. Controlling for age differences between Welsh- and English-medium groups, no significant group differences were found for Reading Aloud or Writing. Welsh-medium children showed lower estimation errors overall and on the 0–100 task (significant/borderline), but not on the 0–20 task.	The results showed two language-specific differences: (1) the estimates of the Italian children were more accurate overall, despite being significantly younger, and (2) the Italian children performed particularly better when inversion errors led to large estimation errors in German children. Italian did not make errors like guessing 48 when the target number was 84, but the Austrian children did frequently.
Additional observations	Pre-intervention, Turkish-speaking children showed lower rote counting and general numeracy skills than English-speaking peers despite Turkish's more regular number naming system, contrasting with prior US–Asian findings. Unlike earlier cross-cultural studies where cardinality performance was similar across groups, Turkish children also underperformed on cardinality tasks. Results suggest that number naming regularity alone does not confer advantage without sufficient numeracy learning experiences. Lower performance was attributed to limited emphasis on mathematics in Turkish kindergartens compared to greater early numeracy exposure in Canada. Although Turkish children started lower on most	Predictors & Relationships: Accurate two-digit (incl. teens) number writing was associated with stronger central executive skills in both language groups; higher non-verbal reasoning additionally predicted accuracy in German-speaking children. Inversion errors were more frequent in children with lower central executive skills (both groups), lower non-verbal reasoning (German), and lower visuo-spatial working memory (English). Arithmetic performance was predicted by non-verbal reasoning and central executive in both groups; verbal and visuo-spatial working memory additionally predicted arithmetic in English speakers. No significant cross-language differences were found in cognitive predictors of arithmetic.	The hypothesis that reading and writing two-digit numbers would be affected by the language group was not supported at all. It may be that these skills, which are school-taught, depend more on specific teaching, which would be similar in the Welsh-and English-medium schools, and do not depend strongly on internal numerical representations. The results imply that greater transparency in a language's counting system may lead to the developmental shift in children's mental number line representation (that is, the shift from a logarithmic representation to a linear representation) occurring at an earlier age	Italian- and German-speaking children share similar underlying mental number line processes, suggesting base-10 understanding is language-invariant. However, German-speaking children showed worse two-digit estimation due to inversion in number words, indicating language influences execution even when underlying processes are shared.
Conclusion	Draw the conclusion that , the results of the study "demonstrate that young children who are given systematic exposure to the counting string learn more quickly if their counting system is regular." (p. 242) but that "[t]he overall pattern of findings suggests that numerical development may be affected more by the quantity or quality of children's early numeracy experiences than by the structure of the number naming system." (p. 243)	The researchers found that German-speaking children made more inversion errors for numbers larger than 20 than English-speaking children. In addition, and for the first time, they reported inversion errors in English-speaking children. These errors occurred more often for number words that are inverted at the item level, i.e., teen numbers in English. Our study shows that an inverted spoken number word leads to more inversion errors, even in a language where most number words are not inverted. They also found that two-digit number writing is a unique concurrent predictor of arithmetic skills in German and in English, showing that the relationship between transcoding ability and arithmetic is not specifically driven by number word inversion.	No differences were found in overall arithmetic or transcoding between Welsh- and English-medium pupils, but Welsh-medium pupils showed better non-verbal number line estimation, particularly for numbers over 20, with smaller error ranges. This advantage is attributed to the transparency of the Welsh counting system, despite similar school environments and curricula, and most Welsh-medium children speaking English at home. The study suggests that a regular counting system supports both number word–symbol correspondence and non-verbal numerical magnitude development.	The findings suggest that the characteristics of a child's native language, such as the counting system and cultural practices, influence their mental representation of numerical magnitudes. Children's nonverbal number line estimations are shaped by the structure and conventions of the counting system in their respective languages.

Recommendations	Researchers propose that differences in counting practices and exposure to numeracy activities within the respective cultures may contribute to the observed variations in overall abilities. Furthermore, learning games and experiences need to be tailored to different languages and number naming systems, e.g. with children speaking languages with less transparent number naming systems needing more frequent exposure to linear numerical board games like the one used in this study.	None	none	None

Details	Study #5	Study #6	Study #7	Study #8
Author (Year)	**Lê & Noël (2020)**	**Li et. al (2013)**	**Magargee & Beauford (2016)**	**Villarroel et al. (2011)**
Citation	Lê, T. P., & Noël, M.-P. (2020). Transparent number-naming system gives only limited advantage for preschooler's numerical development: Comparisons of Vietnamese and French-speaking children. PLOS ONE, 15(12), e0243472.	Li, X., Sun, Y., Baroody, A. J., & Purpura, D. J. (2013). The effect of language on Chinese and American 2- and 3-year olds' small number identification. European Journal of Psychology of Education, 28, 1525–1542.	Magargee, J., & Beauford, K. (2016). Do explicit number names accelerate pre-kindergarteners' numeracy and place value acquisition? Educational Studies in Mathematics, 92, 179–192.	Villarroel, J. D., & Miñón, M. (2011). The origin of counting: A study of the early meaning of "one," "two," and "three" among Basque- and Spanish-speaking children. Educational Studies in Mathematics, 76, 345–361.
Abstract	Number-naming systems vary across languages in terms of transparency, which may influence early numerical development. This study compared preschool children speaking Vietnamese, a language with a highly transparent number-naming system, and French, which is less transparent. Children were tested on counting, number comparison, and simple arithmetic tasks. Results showed that while Vietnamese-speaking children performed slightly better on some counting tasks, the advantage of a transparent number-naming system was limited and did not generalize across all numerical skills. These findings indicate that linguistic transparency provides only modest support for early numerical development, and other factors such as educational experiences also play an important role.	Recent research has found that linguistic cues may affect children's number word acquisition. Two studies were undertaken to evaluate the use of singular/plural markings and small number words in Chinese and English and its effect on children's number concepts. The first study utilized the CHILDES data and investigated how singular/plural markings and number words of "one," "two," and "three" are used in Chinese and compared it with English. The second study entailed testing twenty-four 2.4- to 3.3-year-old Chinese- and English-speaking children on number identification tasks. Results show that the English-speaking children performed significantly better than Chinese-speaking children in identifying "two," "three," and "four" as referents for collections of multiple items but not in identifying those number words as referents for specific pluralities. Additionally, the results show that English-speaking children were more inclined to count when quantifying small number sets than	The purpose of this longitudinal study is to investigate whether an early childhood intervention using an explicit and transparent number naming system will have a lasting benefit to English and Spanish speaking children in their mathematics achievement related to number sense by accelerating their acquisition of concepts of numeracy and place value recognition. Statistical analysis of the results of the Stanford Achievement Test administered to school children in grades 1 through 5 at a school in Texas (USA) over a six-year period showed consistent differences between children who had participated in the intervention program learning to count using explicit number names in Pre-Kindergarten and those who had not participated in the program. The advantages of the study were most apparent in children in the fourth grade, six years after their Pre-Kindergarten experience with explicit number names, particularly for those who had two years of experience with explicit number names	This study examines the beginning of the conceptual understanding of the first number-words and what role language can play in developing the notion of numbers. To that end, 2½-and 3½-year-old Basque and Spanish monolingual children's (N=131) basic numeracy skills are analysed by means of two different experimental procedures: Give-N (in which children are requested to gather sets of objects) and How-Many (based on the ability to count collections). The paper accounts for differences as regards the performance in the Give-N procedure between children with different linguistic backgrounds. In accordance with previous research, this finding can be related to the dissimilar ways by which languages (Basque and Spanish, in this case) express grammatical number, supporting the idea that language plays a definite role in the emergence of the earliest set-size meanings of 'one,' 'two,' and 'three' number-words. Eventually, the work attempts to contribute to the growing body of evidence that shows that the meaning of
Methodology	Eight numerical tasks and general cognitive abilities test + parent questionnaire	CHILDES data base analysis and number word identification tasks - point-to-n and how-many	Longitudinal study over 6 years starting pre- Kindergarten	Give-N and How-Many Tests
Purpose	Explore multiple factors simultaneously or attempted to control or match groups on some variables; for example, the role of the number naming system and parents' influence simultaneously	Evaluate how use of sing/pl markers and small number words affect children's NW acquisition and number concepts.	Investigate whether an early childhood intervention using an explicit and transparent number naming system will have a lasting benefit to English and Spanish speaking children in their mathematics achievement related to number sense by accelerating their acquisition of concepts of numeracy and place value recognition	Examine beginning of conceptional understanding of first number words and role of language in developing of notion of numbers

Research Questions	To what extent does a language's number-naming system impact pre-schoolers' numerical development?	Study 1: 1: Do singular/plural markings exist in Chinese? If so, to what extent, are the singular/plural markings used? 2: How are the number words of "one," "two," and "three" used in Chinese? Study 2: 1: Do English-speaking children identify singular instances as "one" and pluralities with other number words better than Chinese-speaking children? 2: Do singular/plural markings interfere with English-speaking children's appropriate use of "two?"	Does early childhood intervention using explicit and transparent NNS have lasting effect on maths achievement related to number sense by accelerating acquisition of concepts of numeracy and place value?	Does exposure to a greater frequency of linguistic plural markers linked to number-words result in earlier discrimination between number 1 (singular) and the other number-words (plural) and, therefore, an earlier grasp of the first level of numerical comprehension?

Details	Study #5	Study #6	Study #7	Study #8
Hypotheses	H1: We hypothesized that Vietnamese pre-schoolers would perform better on acquiring counting sequences, enumeration, and simple addition than their Belgian peers. For counting ability, we did not expect differences for numbers up to 10, but for larger numbers when number word structures differ. However, based on the lack of the singular/plural distinction in Vietnamese, we expected that Vietnamese children's understanding of number words' cardinal value would be less than that of Belgian children. H2: Furthermore, we assumed that cross-cultural differences in home numeracy would also contribute to the differences between the two samples in the numerical tasks.	Study 1: none Study 2: 1: These markings should only directly affect children's differentiation of singularities from pluralities and the correct use of "one." Therefore, in contrast with Chinese-speaking children, we predicted that English-speaking children should demonstrate superior performance in recognizing "two," "three," and "four" as referents for collections of more than one, but may not demonstrate superior performance in identifying "two," "three," or "four" with a specific plurality. 2: Chinese-speaking children between 25 and 39 months should not experience interference due to learning the add-s rule and—unlike English-speaking children of a comparable age—should continue to use "two" to describe pairs of objects exclusively.	They assume that the earliest possible introduction of explicit number names would yield the greatest advantages.	The frequency of children who have grasped the difference between one and more than one (one-knower level) will be lower among Basque-speaking children, whose language does not refer to the multiplicity of objects with plural markers. However, the overall proportion of subset-knower in both groups will be related to age but not to sex.
Languages	Vietnamese, French	English, Chinese	English, Spanish, expl. number names modelled after Mandarin language	Spanish, Basque
Assigned condition	language groups by native country/language	language groups by native country/language	Those children in the same school system and grade levels who had not attended the school during their Pre-Kindergarten or Kindergarten years and had not received the explicit number names intervention served as the comparison group. The comparison group of children also differed from the intervention group in having attended a variety of other schools and educational settings, thereby experiencing a diversity of other forms of instruction and curriculum	language groups by native language/language of instruction
Mono-/Bilingual	Either monolingual Vietnamese speakers or monolingual Belgian French speakers.	not specified	not specified	All the children included in the Basque-speaking group resided in the Basque Autonomous Community, had Basque-speaking parents, and attended Basque medium schools. Correspondingly, the Spanish sample was exclusively comprised of children who resided in Communities in which Spanish is the official language while Basque is neither an official nor spoken language, whose parents were Spanish native speakers, and who attended infant schools in which Spanish was the sole educational language.
Setting	Kindergarten	Preschool and daycare	Pre-Kindergarten and kindergaten with testing from grade 2 through 6	Infant schools

Country	Vietnam and Belgium (French speaking part)	Chinese data in CHILDES + China and USA	Texas, USA	Within and outside Basque Autonomous Community, Spain
Number Naming Variations	Very regular + base 10 explicit (Vietnamese) vs irregular (French) NNS	Plural markers (English) and no plural markers and different NWs for "two" depending on context counting/cardinality/cardinality designating people (Chinese)	Explicit and transparent vs English and Spanish	Plural markers (Spanish) and no plural markers (Basque) assoc. with cardinal numbers

Details	Study #5	Study #6	Study #7	Study #8
Setup/ design	The general and numerical development of the children was assessed through General cognitive tasks, Numerical tasks, and Parental home activity questionnaire	The participants were assessed on their small number identification skills by matching small sets of objects (ranging from 1 to 4 items) to corresponding number words. For this assessment the point-to-n task and the how-many task were used.	The Mandarin-based number naming system used in this study is structured as followed: The symbolic representations are read from left to right in descending order of place value. For example, a quantity of 17 named "one-ten seven" not only explicitly names the exact amount but also identifies the place values and the exact amounts within these place values. Children learn the number families of digits 10 to 19, 20 to 29, 30 to 39, and so on, by explicit statement of the number of tens in each family, 'one-ten', 'two-ten', and 'three-ten'. Thus each family is stated as follows: 0 to 9 as 'ones', 10 to 19 as 'one-ten^' 20 to 29 as 'two-ten' and so on, up through 99.	The researchers presented the children with two different tasks, the Give-N and the How-Many tasks. For the Give-N test children are presented with a set of objects and are required to select a subset of a given quantity from it. After having completed the Give-N task, every child was shown a set of the same plastic fruit used previously and asked to count how many fruits there were in the set.
Pre-/post-test	not applicable	not applicable	For assessment the Stanford Achievement Test was used, which measures student proficiency in basic grade level mathematic concepts and skills while providing a comparison to both local and national peer performance.	not applicable
Testing procedure	Cognitive and numerical assessments: Non-verbal intelligence (Matrix Reasoning, 4+ yrs) and phonological loop (letter repetition). Numerical tasks: counting, advanced counting, enumeration, Give-N, number-word and collection comparison, addition, and approximate addition. Parents completed a home numeracy/literacy questionnaire. Testing: 2 individual sessions per child (3.5–4 yrs: ~25 min; 4+ yrs: ~45 min), younger children skipped matrix and addition tasks.	Children were tested individually in four 15-minute sessions. The Point-to-N task used real small replicas to assess whether children could identify sets matching a number word (2–4), with two sessions of four rounds per number. The How-Many task measured children's ability to report the cardinality of item sets (2–4) within story and game contexts, using pictures or toys, with multiple trials per number.	Teachers in the explicit number names group received targeted training before each school year, emphasizing teaching explicit number names and corresponding quantities. In Pre-Kindergarten, explicit number names were used exclusively for introducing numbers, with teachers applying them consistently in counting, measuring, telling time, and dates—this was the only curricular change. The traditional instruction group followed the regular math curriculum without emphasis on explicit number names. Students were expected to rote count to 100 by Kindergarten end; thus, explicit names were used exclusively for the first three months, after which traditional number names were introduced and blended with explicit names for the rest of the year. Teachers used both naming systems interchangeably thereafter. In Spanish classes, children were taught to count using an explicit number naming system similar to Mandarin-based systems, but with 'one-ten one, one-ten two…' used for teens instead of 'ten one, ten two…'.	Give-N: Experimenters asked the child to help them select a certain number of plastic fruits from an array of fruit. When a child selected the correct number n, they were then asked to select n+1 fruit in the next attempt. If the child was unable to select n+1 fruit, they were then asked to select n fruit again during the following test. Each child carried out three rounds of testing made up of five tests each. Each round started with the child being asked to select one single fruit and then to either select two or three, alternating from round to round, and finally to select four, five or more fruit. How-many: After having completed the Give-N task, three rounds of How-many- testing were administered, in which the child was presented with sets of 1, 2, 3, 4, or more items. A response was scored as correct when a child used the specific, correct cardinal number corresponding to the overall number of items in a given set.
Scoring	not applicable	Point-to-n task: A response was considered correct if a child pointed to the correct mat without counting. How-many task: A child's performance was counted as correct, if they spontaneously labelled a set, or enumerated the set and then labelled it correctly.	Achievement Testing: Longitudinal data collected via the Stanford Achievement Test from Grade 1 onward. Intervention group initially received explicit number names; comparison group received traditional instruction. Assessed domains: number sense & operations, algebra, data/statistics/probability, geometry/measurement, estimation, reasoning/problem solving, mathematical connections, computation (whole numbers, decimals, fractions, symbolic notation), and thinking skills. Scores reported for Grades 1–5 for both groups.	Give-N: Children categorized by counting ability: no-number-knower (cannot select any subset), one-/two-/three-knower (can select subsets of 1–3), CP-knower (can select subsets of ≥4). Maximum one error allowed per level. How-Many: Correct responses included verbally stating the number, counting aloud then stopping at the correct number, switching from aloud to silent counting, or showing the correct number with fingers.

Details	Study #5	Study #6	Study #7	Study #8
Key findings	Non-verbal IQ and phonological loop increased with age, similar across language groups. Age positively influenced all numerical tasks; counting showed a language effect (31 Belgian vs. 3 Vietnamese children stopped at teen level). Home numeracy activities were more frequent than literacy and higher in Vietnamese families, but children's symbolic numerical skills were similar except for basic counting. Home numeracy correlated with precise numerical tasks but not approximate tasks, consistent across groups.	Results (Point-to-N & How-Many): In the Point-to-N task, Chinese children made errors distinguishing singular vs. plural sets (1 vs. 2/3/4), whereas American children made none. In the How-Many task, American children significantly outperformed Chinese children on "two," marginally on "three," with similar (non-significant) trends for "four." Chinese children significantly over-applied "two" to larger sets more often than Americans. Both groups mislabelled sets, but Chinese children more frequently misapplied number words (including labelling "two" as "one"). American children engaged in spontaneous counting more often and with higher accuracy across set sizes (2–4 items).	Children receiving early, explicit number names instruction in Pre-K showed the greatest gains in numeracy, counting, cardinality, number comparison, and place value, with strongest effects in Number Sense and Calculation with Decimals. Advantages persisted through 4th grade, indicating long-term benefits of explicit number name teaching over traditional instruction.	Regarding the relationship between language profile and counting skills in there was no difference between the tested Spanish- and Basque-speaking children in their ability to use number-words for counting (How-many task). When it comes to correctly label the cardinality of subsets of items, there was a significant difference between the language groups in the understanding of set-size meanings of the number-words 'one,' 'two' and 'three'. The frequency of Basque-speaking children who do not display the most elementary level of numerical understanding (no-number-knower) as well as those who are not able to distinguish sets of 'one' from sets of other numbers (one-knower) is significantly higher. Both groups of children had similar understanding of 'two', 'three', and higher numbers when referring to sets of
Additional observations	For the Give-N task, the language group effect was only significant on the two-knower level, indicating that more French-speaking children could understand the cardinal meaning of "two" than Vietnamese-speaking children. (Singular/Plural distinction.) The analysis of the two additional tasks performed by the older children only also showed a significant age effect, performance improved with age for both addition and approximate addition, but no significant language group effect could be found for either task.	Results (Point-to-N & How-Many): In the Point-to-N task, Chinese children made errors distinguishing singular vs. plural sets (1 vs. 2/3/4), whereas American children made none. In the How-Many task, American children significantly outperformed Chinese children on "two," marginally on "three," with similar (non-significant) trends for "four." Chinese children significantly over-applied "two" to larger sets more often than Americans. Both groups mislabelled sets, but Chinese children more frequently misapplied number words (including labelling "two" as "one"). American children engaged in spontaneous counting more often and with higher accuracy across set sizes (2–4 items).	Even after Kindergarten, children who were part of the intervention group kept using explicit number names in their problem solving sometimes, even though explicit number names were not mentioned anymore in first grade and subsequent classes.	No significant differences between boys and girls in the sample were found regarding verbal counting (How-Many task) or the ability to correctly select subset (Give-N task). Additionally, there has been no relationship found between age and the level of set-knowers (Give–N task) but there was a significant age effect on the levels of verbal counting abilities (How-Many task) in that the older the children, the higher their number-knower level tended to be.
Conclusion	Vietnamese-speaking children outperformed French-speaking peers only in counting. Results suggest that both parent-child numeracy stimulation and number-naming transparency influence counting, but transparency alone did not confer an advantage in other numerical tasks. Cultural and educational experiences likely play a larger role in early numerical development.	Chinese children outperformed American children in small set identification and matching sets to numerals/number words. However, findings indicate that singular/plural marking supports early number word acquisition: while it does not directly teach the meaning of "two," it helps children distinguish "two" from "one." In how-many and point-to-n tasks, American children performed better in identifying "two," "three," and "four" as non-singular quantities, with Chinese children making significantly more errors and more frequently associating these numbers with singular sets. Chinese preschoolers were also more likely to over-apply "two," suggesting that overgeneralisation of number words is not unique to English-speaking children.	The study highlights the importance of focusing on explicit number names in early mathematics instruction. By teaching explicit number names and their corresponding quantities, educators can provide a solid foundation for children's numerical development and enhance their ability to work with numbers effectively.	Language and cultural factors influence early number-word development. Basque lacks singular/plural markers, unlike Spanish, which affects children's understanding of the cardinal meaning of small numbers. Spanish-speaking children outperformed Basque-speaking peers on the Number-knower task, especially in distinguishing 'one' from other numbers, though both groups performed similarly on counting tasks. The difference is linked to Basque's absence of plural marking in nouns and adjectives.

Recommendations	Overall, the study provides insights into the role of number-naming systems in pre-schooler's numerical development. It challenges the notion that a transparent number-naming system alone leads to a significant advantage in numerical skills. The findings emphasise the need to consider a broader range of factors when studying the development of numerical abilities in children, beyond the transparency of the number-naming system.	Early advantages in small number word learning for American children, linked to singular/plural marking, fade over time. Later, Chinese children outperform Americans in broader mathematical tasks, likely due to the highly regular Asian counting sequences, which facilitate learning larger counting sequences compared to the irregular English counting rules for teens and decades.	Overall, the study provides evidence that explicit instruction in number names can accelerate pre-kindergarten children's numeracy and place value acquisition. The findings have implications for early mathematics education, emphasizing the significance of teach explicit number names to support children's development of numerical concepts and place value understanding.	none

Appendix 3. Analysis of Individual Studies located in the Scoping Review

Study #1, Cankaya et al (2014)

Study #1, essential characteristics

Purpose	To investigate how diverse number naming systems and numeracy experiences affect young children's development of rote counting and other early numeracy skills
Languages	Turkish and English
Setting	Kindergarten and day-care
Number naming variations	Regular (Turkish) vs less regular (English) number naming systems
Methodology	Number game involving counting and naming numbers
Participants	3 -4.5-year-old children; the study only included subjects who were not able to count higher than 20 without error during pre-test

Study #1, participants

Countries and Languages	Overall	Female	Male
Turkey (Turkish-speaking)	45	19	26
Canada (English-Speaking)	30	16	14

The study by Cankaya et al. (2014) investigated the influence of number naming systems and numeracy experiences on children's acquisition of numeracy skills, especially rote counting abilities, focusing on different levels of transparency of the Turkish and English number naming systems, as well as wider cultural variations in numeracy experiences between Turkish and Canadian children.
The research question was: "How important is the regularity of a number naming system when young children are learning to count?" (p.238). The study's hypotheses were:

- A relatively regular number naming system (Turkish) will facilitate children's acquisition of counting words more than an irregular number naming system (English) when children who did not have much prior counting knowledge

were exposed systematically to number names and the counting string.
- Exposure to number games will boost early numeracy skills for both English- and Turkish-speaking children, relative to exposure to colour games.

The experimental set-up was unique among the studies (a number game) and the results showed connections between transparency of the number names (Turkish children profiting more from a short intervention and showing faster processing of simpler system). The findings also show the impact of outside influences as English children benefited from earlier numerical education as they scored higher than the Turkish children before the intervention. The experimental design in this study involved a game with either a bridge built of squares in the recurring colour scheme of red, yellow, green, and blue or the digits 1 through 20 printed inside the bridge's squares. Spinners in the colours of yellow, red, green, and blue or the numerals 1 and 2 twice each inside the spinner square, respectively, were placed next to the board. Roughly half of the children in each language group were assigned to either the colour game or the number game condition. The children were given the option of moving their token, depending on which square the spinner landed on, to the colour indicated or either one or two squares ahead. The children were then asked to name the colour or number found inside the square that the token landed on. To assess the effects of the intervention on rote counting performance children were given six attempts to rote count as high as they could during each pre-test, four game sessions, and post-test. For each session the highest count the child was able to achieve out of the six attempts was determined and compared with the highest count from every other session and the resulting data was analysed to explore patterns of change.

Since counting performances were highly variable even at post-test level, the researchers deemed this method of scoring to be a better index to reveal performance patterns as opposed to only taking the pre-and post-test scores into consideration. Then they determined the number of children with the highest counts in each stage across language and game groups. The stages were as follows: before intervention (pre-test or before first game-session), during intervention (before the second or third game-session), and after intervention (before the last game-session or post-test). A higher percentage of children in the later stages would indicate that the intervention was successful in facilitating an improvement in rote counting ability.

The results showed that more Turkish-speaking children benefitted from the training more than English-speaking children. They achieved their highest counts after the intervention (before 11%, during 36%), which agrees with the first hypothesis. In contrast, more English-speaking children achieved their highest count before the intervention (b. 47%, d. 20%, a. 33%). Additionally, improvement depended on game condition, with more children who played the number game during training showing improvement after the intervention (b. 18%, d. 16%, a. 66%). Furthermore, all Turkish-speaking children showed improvement, while only half of the English-speaking ones did. The children from both language groups who had been assigned to the colour game condition did not show improvement (b. 32%, d. 43%, a. 24%) however.

Study #2, Clayton et al (2020)

Study #2, essential characteristics

Languages	German and English
Setting	Primary school
Number naming variations	Fully-inverted (German) vs minimally-inverted (English, teen numbers 13 – 19 only)
Participants	Children from the UK, mean age of 6 years 2 months and from Austria, mean age 7 years 1 month.

Study #2, participants

Countries and Languages	Overall	Female	Male	Mean Age
UK (English-speaking)	309	142	167	6.2 years
Austria (German-speaking)	177	83	94	7.1 years

Clayton et al. (2020) examined the impact of number word inversion on the two-digit number writing and arithmetic abilities of 177 German-speaking and 309 English-speaking Year 1 children attending primary schools in Graz, Austria and Yorkshire, UK, respectively. The focus was on Year 1 children, because they are in the early stages of learning two-digit numbers and arithmetic concepts. The mean age of the Austrian children was 7 years and 1 month and the mean age of the British ones was 6 years and 2 months. The researchers investigated whether the inversion of number words, such as saying "twenty-three" in English instead of "three-twenty", the way all two-digit number words are structured in German, affects children's numerical skills. Special focus of this study was on transcoding errors in teen numbers, since teen number words in both German and English consist of two types: particulars (e.g., 'eleven' in English, elf in German) and combinations of a single-digit number word with a teen marker (e.g., 'sixteen' in English, *sechzehn* in German). Additionally, the latter kind are the only two-digit number words in English that are also inverted. The stated purpose of the study was to:

- Provide more information about how language affects the transcoding of two-digit numbers
- Conduct a more thorough analysis of the connection between number writing and arithmetic. To compare specifically the relative contributions of two-digit number writing in German with inverted number words and English with non-inverted number words as a concurrent predictor of arithmetic abilities.

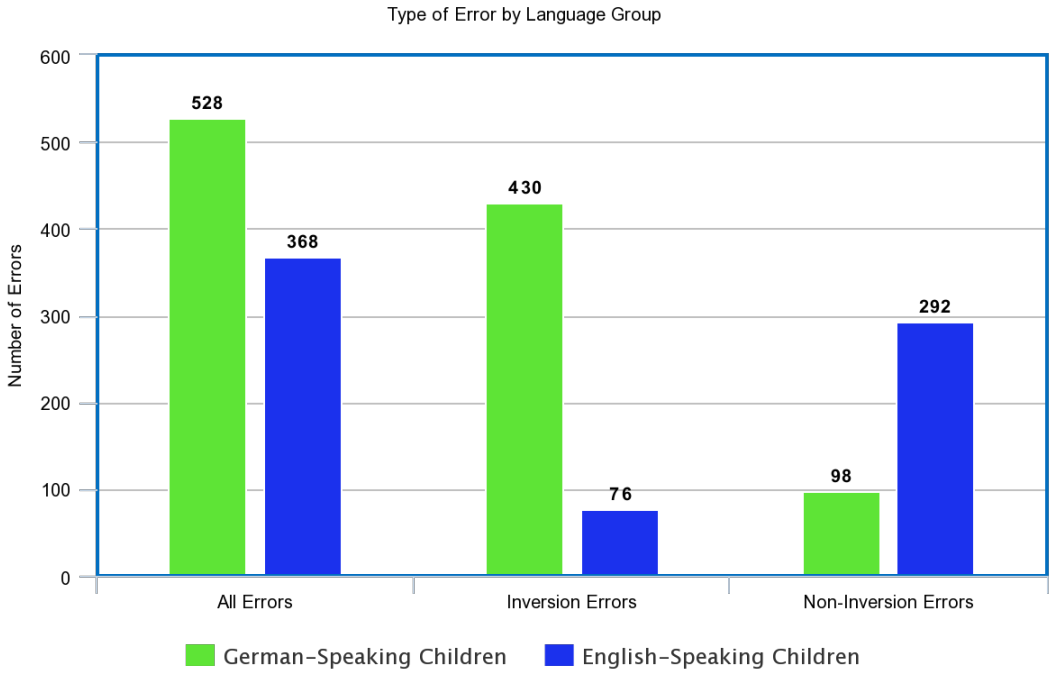
The following research question was posed: "Does number word inversion lead to a general increase in transcoding errors, i.e., also in non-inversion errors, or does it specifically affect only the number of inversion errors?" (Clayton et al, 2020, p. 2). The predicted hypothetical outcome was that, due to the inverted structure of teen number words in both languages, German-and English-speaking children would make inversion errors. However, it was hypothesised that German-speaking children would make more inversion errors for two-digit numbers larger than 20 and fewer non-

inversion errors for two-digit numbers larger than 20 than English-speaking children. Researchers hypothesised that performance in arithmetic would be significantly predicted by the number writing ability for two-digit numbers in both languages. The focus on inversion errors, taking into account the fact that English has a very limited amount of inverted number names, makes this a unique study. The effect of inversion on numerical processing in children is demonstrated much more persuasively when there is only a limited scope of inverted number names, like in number thirteen to nineteen in English, and inversion errors occur The study showed that English-speaking children made more inversion errors for teen numbers only, whereas German-speaking children made more inversion errors overall.

The participants were assessed on four tasks: number writing, arithmetic and non-verbal reasoning in a classroom setting, and working memory was assessed individually. Four assessed tasks were part of a larger battery of measures administered in a fixed order in a classroom setting in two one-hour sessions. Working memory was tested individually. All tests were administered in English for the English-speaking and in German for the German-speaking children. The four tasks were number writing, arithmetic (Wechsler Individual Achievement Test, non-verbal reasoning (Ravens Standard Progressive Matrices Plus), and working memory: (Working Memory Test Battery for Children).

The key findings of the study suggest that, across language groups, the overall occurrence of inversion errors as a proportion of all errors made in writing two-digit numbers differed significantly (Figure7.2.). There were 430 inversion errors (81%) compared to 98 noninversion errors (19%) in the German-speaking group, whereas in the English-speaking group there were 76 inversion errors (21%) compared to 292 non-inversion errors (79%). The German-speaking children had a higher percentage of inversion errors for two-digit numbers than for teens and a higher percentage of non-inversion errors for teens than for higher two-digit numbers. The English-speaking group showed the opposite pattern; children had a higher percentage of inversion errors for teens than higher two-digit numbers and a higher percentage of non-inversion errors for two-digit numbers than teens.

Study #2, Proportions of errors made by language group 2



Language group	All errors	Inversion errors	Non-inversion errors
German-speaking children	528	430 (81% of all errors)	98 (19%)
English-speaking children	368	76 (21%)	292 (79%)

Study #3, Dowker & Roberts (2015)

Study #3, essential characteristics

Methodology	Nonverbal MNL estimation and transcoding
Languages	English and Welsh
Setting	Primary school
Number naming variations	Regular (Welsh) vs less regular (English) number naming systems´.

Study #3, participants

Languages	Overall	Female	Male	Mean Age
English-Medium	20	14	6	6.7 years
Welsh-Medium	20	10	10	6.5 years

Dowker and Roberts (2015) explored the impact of the transparency of the counting system on children's numerical abilities. The researchers investigate whether the ease of mapping between number words and quantities in a counting system affects children's performance in numerical tasks. There are few studies available using a language other than Chinese as the transparent example, and the present study is the only one using Welsh. The researchers tested 40 6 and 8-year-old Year 2 pupils, 20 of whom attended Welsh-medium schools and 20 English-medium schools in Wales. In this way, the researchers were able to bypass the problem of different cultures and educational systems. The children were tested with a variety of number processing tasks.

Study #3 focused on the transparency of counting systems, which refers to the extent to
which the number words reflect the base-10 structure of the numerical system explored earlier in Chapter 3. For example, in English, the number words from 11 to 19 have unique names that do not follow a regular pattern. In contrast, the Welsh language contains two counting systems, an irregular one and a regular one, and only the latter is used to teach arithmetic in schools (Roberts, 2000). This counting system is transparent, 11 is un deg un, which literally translates to “one-ten-one” and 22 is dau degg dau or “two-tens-two”. As in Chinese and Japanese, Welsh number words have a transparent structure, with a consistent pattern for numbers from 11 to 99. Study #3’s purpose was to explore whether linguistic qualities such as transparency affect how well children perform on specific arithmetic problems. To this end, the following research question was addressed in this study: “Does the language of school instruction affect the mental number line in children who are otherwise having similar cultural and educational experiences?” (p.3). The researchers hypothesised that:

- H1: Children in the Welsh-medium system would score higher on the number line estimation tasks, and this would especially be true of the number line from 1 to 100, as this places greater demands on the ability to represent multi-digit numbers, for which a transparent counting system would provide an advantage.

- H2: Children in the Welsh-medium system would perform better on the tasks involving reading and writing numbers, but there would be no difference between the two groups in BAS Number Skills test performance (based on Dowker et al.’s (2008) earlier findings and their suggestion that linguistic

effects on mathematics may be specific rather than global).

This study took advantage of the fact that researchers could study two different number naming systems within the same educational context and the same cultural background. In this regard the results were clearer cut than similar studies conducted in different countries and cultural environments. However, the researchers did not explicitly take into account the fact that all children enrolled the Welsh-medium schools were bilingual and also spoke English as their first or one of their first languages. It is difficult to control and/or measure children's prior knowledge of other languages and number systems, and this is a limitation of this kind of study. The experiment asked the children to complete the following four tasks:

- British Abilities Scales Number Skills test, a standardized test that assesses written calculation (Elliott et al., 1997).
- Two transcoding tasks (Writing to Dictation and Reading Aloud).
- A non-verbal number line estimation task in one-on-one sessions with the experimenter. The difference between the shown number's actual position on the number line and the number that the child estimated it to be, was measured to the nearest millimetre in order to calculate the child's number line estimation scores.

The fact that the language groups differed significantly in age and the fact that age correlates significantly with the Reading Aloud and Writing tasks was considered when carrying out the analysis. However, when comparing the scores of the Welsh- and English-medium groups, the differences between groups did not approach significance for either task. Estimation errors were lower in the Welsh-medium group overall. While there was a significant difference for the overall estimation error score, and a borderline significant difference for estimation errors in the 0–100 number line task, there was no significant group difference for the 0–20 number line estimation error. The latter indicates that the fact that English "teen"-numbers are inverted while Welsh numbers did not affect estimation abilities in either group.

Study #4: Helmreich et al. (2011)

Study #4, essential characteristics

Purpose	Investigate if language properties, namely inversion, influence children's performance in non-verbal number line estimation tasks
Methodology	Nonverbal MNL estimation and general cognitive abilities test
Languages	German and Italian
Setting	Primary schools in Austria and Italy
Number naming variations	Fully-inverted (German) vs minimally-inverted (Italian)

Study #4, participants

Countries and Languages	Overall	Female	Male	Mean Age
Austria (German-speaking)	130	63	67	7.4 years
Italy (Italian-speaking)	107	55	52	6.11 years

Helmreich et al. (2011) examined counting systems in German (Austrian children) and Italian (Italian children), which are mostly transparent, but not to the same degree as the Chinese and Welsh counting systems. The potentially most confusing property of the German number naming system is inversion, akin to English "teen"-numbers, for instance, 19 = "nine" (9) - "teen" (1). In German, all tens and units up to 99 are inverted. For example, 21 is ein-und-zwanzig, which literally translates to "one-and-twenty and 56 is sechs-und-fünfzig or six-and-fifty.Italian uses inversion for a section of the Teen-class, namely for the counting terms 11-16, un-dici (eleven), do-dici (twelve), tre-dici (thirteen), quattor-dici (fourteen), quin-dici (fifteen), and se-dici (sixteen). Counting words from seventeen upwards are in standard order, but sometimes do use the and-infix, compare dici-as-sette (ten-and-seven) with dici-otto (ten-eight).

These examples show that departures from a counting sequence's general structure are a common occurrence in languages. What matters is that this is limited to a few locations and a small number of elements, so these sub-sequences or individual elements may be learned as distinctive. The study focused on children's nonverbal number line estimations, which involve placing numbers on a mental number line based on their perceived magnitudes.

The researchers aimed to understand how language factors, i.e. inversion of number words, may shape children's numerical representations and estimations. They posed the following research question: Does language as reflected by the structure of its number word system influence performance in the non-verbal spatial-numerical number line estimation task? (Helmreich et al., 2011, p. 600). They hypothesised that:

- H1: Austrian children's estimating inaccuracy would rise due to the German tens and unit order being reversed.
- H2: In addition to the log-to-linear assumption, a two-linear model that suggests that mastering the Arabic number system's place-value structure is necessary for forming a mental representation of number size may also explain children's overall estimating performance.

- H3: In the context of place-value integration, German-speaking children's estimations should be especially inaccurate for numbers where mixing up tens and units leads in a considerable estimation error (for example, 82, 28) compared to numbers where this error is just little (for example, 54 45).
- H3A: In connection with this, the following even more specific hypothesis might be made, German-speaking children should particularly underestimate numbers like 82 because doing so leads to a lesser number (28). This under-/overestimation should be more pronounced in German-speaking children because there is no inversion in Italian number names for these higher numbers (ibid.).

The experimental design in this study was very simple and consisted of a paper-and-pencil version of a number-to-position task for the number line task. The children were required to indicate their estimate of a given number with a pencil on an empty number line. The line was 10cm in length, the left side was labelled 0 and the right side was labelled 100. To ensure that no verbal number word was involved in the task at any level, presentation or response, the numbers to be estimated were printed in Arabic notation in the middle of the page above the line.

Participants were tested in one-to-one single sessions with the experimenter. In order to acquire estimation accuracy, the individual deviation of the estimated from the actual value was determined for each child independently after the automatic measurement to the nearest millimetre of the estimated location from the true position for every single item. Results showed two language-specific differences: (i) The estimates of the Italian-speaking children were more accurate overall, despite being significantly younger and (ii) Italian-speaking children performed particularly better when inversion errors led to large estimation errors in German-speaking children. Italian-speaking children did not make errors such as guessing 48 when the target number was 84, whereas the German-speaking children did so frequently.

Study #5: Lê & Noël (2020)

Study #5, essential characteristics

Methodology	Eight numerical tasks and general cognitive abilities test + parent questionnaire
Purpose	Explore multiple factors simultaneously or attempted to control or match groups on some variables; for example, the role of the number naming system and parents' influence simultaneously.
Languages	Vietnamese and French
Number naming variations	Very regular and base-10 explicit (Vietnamese) vs irregular (French) number naming systems
Setting	Kindergarten
Participants	Children aged between 3.5 and 5.5 years from Vietnam and the French-speaking part of Belgium.

Study #5, participants

Countries and Languages	Overall	Female	Male	Mean Age
Vietnam (Vietnamese-speaking)	104	54	50	4.5 years
Belgium (French-speaking)	104	54	50	4.6 years

Lê and Noël (2020) investigated the impact of number-naming systems on the numerical development of preschool-aged children. To explore the impact of transparency in number-naming systems on numerical skills, the study compared Vietnamese children speaking Vietnamese and Belgian children speaking French. French has a less transparent approach than Vietnamese (which is fully regular and base-10 explicit). Additionally, the researchers wanted to investigate how various elements interact simultaneously, such as the role of the number naming system and the influence of parents on at-home numeracy activities. The research question addressed in this study was:

- To what extent does a language's number-naming system impact preschoolers' numerical development?

The following hypotheses were stated:

- H1: Vietnamese pre-schoolers will outperform their Belgian counterparts in learning counting sequences, enumeration, and simple addition, whilst Vietnamese children will have a weaker knowledge of the cardinal value of number terms than Belgian children due to the absence of the singular/plural difference in Vietnamese.

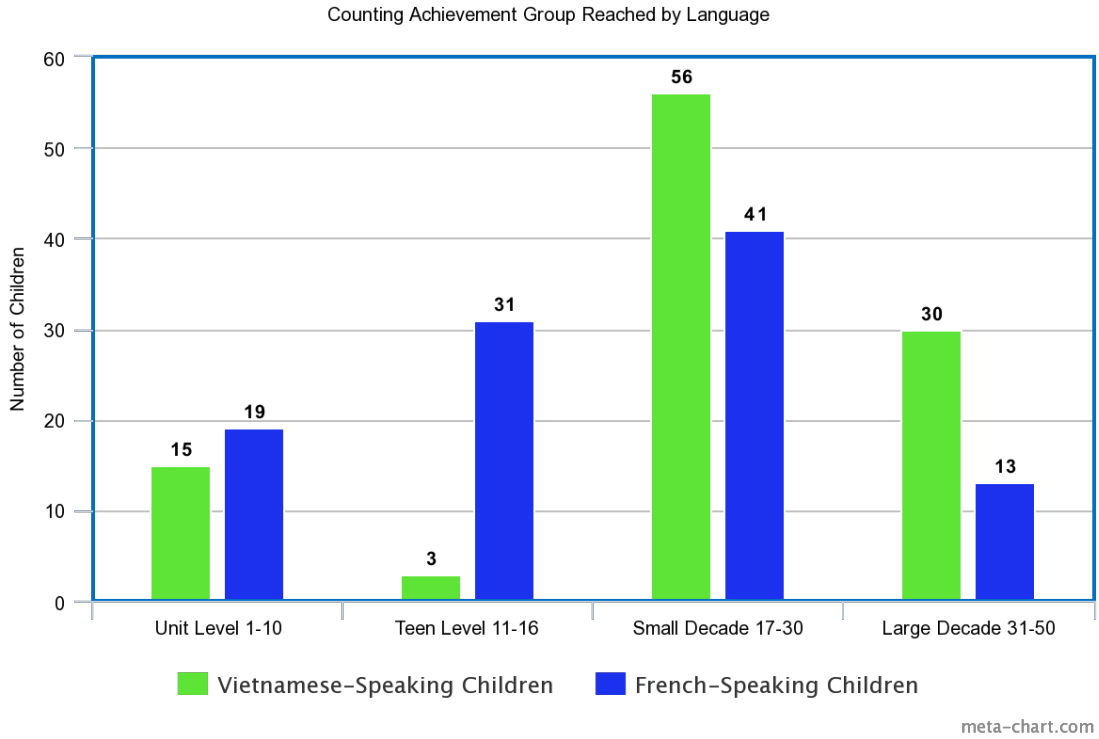
- H2: Discrepancies between the two samples in the numerical tasks will be influenced by cross-cultural variations in home numeracy.

The testing of the children in this study was extensive. The inclusion of outside influences such as home numeracy on early numeric learning is a standout feature of

this study. The researchers also took cultural similarities and differences into account. All of these factors made for a comprehensive study that shed light on the influences of number name systems and parental influences on early numeric learning. The testing materials comprised:

- General cognitive tasks
- Numerical tasks (counting, advanced counting, enumeration, Give-N, collection comparison and number-word comparison, addition, approximate addition)
- Parental home activity questionnaire.

Children were tested individually, and no details were given about the scoring procedures used. All tests demonstrated an increase in performance with age. The counting task showed a notable language group effect as shown below. Only three Vietnamese children were not able to count higher than the teen level (eleven to sixteen). Even though numeracy was stimulated at home slightly more frequently in the Vietnamese context, Vietnamese-speaking children did not show more advanced symbolic numerical development, apart for basic counting skills. Home numeracy did not significantly correlate with the approximate numerical activities, such as collection comparison and approximation addition, but it did significantly correlate with all precise symbolic numerical abilities, such as counting, advanced counting, enumeration, Give-N, number-word comparison, and addition The number-knower results are below. These correlations were unaffected by language group in all instances.



Study #6A and #6B, Li et al. (2013)

Study #6, essential characteristics

Methodology	CHILDES data base analysis and number word identification tasks - point-to-n and how-many
Purpose	Evaluate how use of sing/pl markers and small number words affect children's NW acquisition and number concepts.
Languages	Chinese (Mandarin) and English
Setting	Preschool and day-care
Number naming variations	Number naming variations: Plural markers (English) and no plural markers and different Number words for "two" depending on context counting/cardinality/cardinality designating people (Mandarin)
Participants	2.4 -3.3-year-old children from China and the United States.

Study #6, participants

Countries and Languages	Overall	Female	Male
USA (English-speaking)	12	5	7
China (Mandarin-speaking)	12	6	6

The Li et al. (2013) publication encompassed two studies (labelled here as #6A and #6B for the purposes of this scoping review) conducted consecutively to investigate how linguistic cues may affect young children's number word acquisition. Study #6A investigated the usage of singular/plural markers and number words like 'one', 'two', and 'three' in Chinese and English using the CHILDES database[3]. Study #6B examined the number identification skills among pre-schoolers from China and the United States who were tested to see which effect the different use of singular/plural markers and small number words may have on number word acquisition and identification. The researchers explored how language as well as cultural factors may influence early numerical development in these two populations. Study #6B focuses on 2.4-and 3.3-year-old children as they are in the early stages of developing numerical skills and are beginning to learn to identify small numbers. The research question put forward for each study was as follows:

Study #6A:

- Do singular/plural markings exist in Chinese? If so, to what extent, are the singular/plural markings used?
- How are the number words of "one," "two," and "three" used in Chinese?

Study #6B:

[3] CHILDES is a database of child language data which provides researchers with a collection of corpora of transcripts and related resources, see www.childes.talkbank.org

- Do English-speaking children identify singular instances as “one” and pluralities with other number words better than Chinese-speaking children?
- Do singular/plural markings interfere with English-speaking children’s appropriate use of “two”?

No hypothesis for Study #6A was given. Hypotheses articulated for Study #6B were as follows:

- H1: That English-speaking children should perform better than Chinese-speaking children in recognising “two” “three” and “four” as referents for sets of more than one, but that they might not perform better in associating “two” “three” or “four” with a particular plurality.
- H2: Unlike English-speaking children of a similar age, Chinese-speaking children between the ages of 25 and 39 months are not expected to encounter any difficulties as a result of learning the add-s rule, and they should continue to just use the word “two” to refer to pairs of items.

The set up for testing consisted of the Point-To-N task and the How-Many tasks. The children were tested individually. The Point-to-n task involved the following: If a child pointed to the appropriate mat with objects on it without counting, their response was regarded as correct. In the How-many task, a child's performance was regarded as correct when they correctly labelled a set on their own or counted the set before labelling it.

The findings of study #6A are not discussed here since they are not relevant to the scope of this review. The findings of study #6B suggest that singular/plural markings play a significant role in early numerical development and acquisition of number word concepts. While overall five Chinese-speaking children made mistakes during the Point-To-N task, there were no mistakes made by English-speaking children. Singular/plural markers help children understand that “two” does not refer to just one object, even if they do not specifically facilitate children in understanding “two” as a referent for just two and only two items. The American participants performed better than their Chinese counterparts in both the How-Many and Point-To-N tasks, when it came to identifying "two” “three” or “four” as referents for non-singular quantities.

Study #7 Magargee & Beauford (2016)

Study #7, Essential characteristics

Methodology	Longitudinal study over 6 years starting pre-kindergarten
Participants	4-to 5-year-old pre-kindergarteners and kindergarteners at point of intervention, all children from Texas, USA.
Sample size	101 children, from one pre-kindergarten and one kindergarten class, approximately half female, half male
Languages	English, Spanish, explicit number names modelled after number names in Mandarin Chinese
Setting	Pre-kindergarten and kindergarten
Number naming variations	Instruction in explicit number names

Study #7 examined the effect of explicit number names on the numeracy and place value acquisition of pre-kindergarten children. The researchers aimed to determine whether teaching explicit number names can enhance children's understanding of numerical concepts, such as counting, cardinality, and place value in children coming from language backgrounds with non-explicit number naming systems. The authors investigated whether an early childhood intervention using an explicit and transparent number naming system has a lasting benefit for English- and Spanish-speaking children in their mathematics achievement related to number sense by accelerating their acquisition of concepts of numeracy and place value recognition. The following research questions were posed:

- Is an early introduction of explicit number names to preschool children of non-explicit number naming system heritages beneficial?
- Would any such benefit be long lasting?

The authors hypothesised that the earliest possible introduction of explicit number names would yield the greatest advantages. Study #7 is the only longitudinal study selected for inclusion in this scoping review and it is the study with the largest sample size. The study's design of modelling number names drawn from one language on the model of highly transparent Chinese number names, and then teaching these to children in the process of acquiring early numeracy, is unique amongst the studies identified. The results suggested that early understanding of the base-10 concept has long-lasting benefits. The pre-kindergarteners and kindergarteners were taught transparent systems first before learning the regular ones later. The results of yearly standardised maths achievement tests showed that children who had learned the highly transparent number naming systems the earliest, had the highest scores in all categories tested except one. These advantages were still apparent in the test results six years after their experience with explicitly transparent number names.

To better understand the method and design of the intervention, it is important to understand the nature of the Mandarin-based number naming system used in the study. The Mandarin-based number naming system explicitly encodes place value by reading numbers from left to right, stating the number of tens followed by the ones. For example, 17 is named "one-ten seven," making both the total quantity and its composition clear. Children learn number families in tens (10-19, 20-29, etc.) by this systematic naming, which continues up to 99 (Magargee & Beauford, 2016, p. 18).

The Stanford Achievement Test was administered as the means to test children's mathematical progress over the duration of this long-term study. The test evaluates student proficiency in fundamental grade-level mathematical concepts and skills and provides a comparison to both local and international peer performance. Statistical evaluation of the results was used to distinguish between those children who had learned to count using explicit number names in pre-kindergarten and those who had not. The Stanford Achievement Test's mathematics section is divided into two sub-sections, Mathematics Problem Solving and Mathematics Procedures. Data were divided into three categories depending on how long children had been using explicit number names:

> 0 = pupils who entered the school after kindergarten and who had never used explicit number names.
> 1 = students who entered the school in kindergarten but had only used explicit number names for three months.
> 2 = students who had attended the school since pre-kindergarten and who had used explicit number names since then.

The number of topics varied between grade levels and within subject area, i.e. computation and problem solving. It ranged from 4 to 24 for each topic. Stanford's reporting scheme included some components that were used on several topics. In Grades One through Three, calculations that were dependent specifically on comprehension of number facts were evaluated. At all Grades One through Five, problem solving abilities were evaluated, most notably Number Sense. Naturally, several subjects, most notably fractions and integers, had not been taught yet in the early grades. Stanford reports national and regional rankings for each subject area, but not for specific topics. The percentage of answers that were accurate for each topic was therefore used for further analysis in this study. The results showed that, with the exception of Geometry, the pre-kindergarten students who got the intervention as early as possible, meaning pre-kindergarten at 4 to 5 years of age, had the highest mean scores overall and showed the most notable differences from their classmates in terms of their accelerated learning of numeracy and place value concepts. These advantages were still apparent in children in fourth grade, six years after their early exposure to explicit number names.

Study #8 Villarroel et al. (2011)

Study #8, essential characteristics

Methodology	Give-N and How-Many Tasks
Purpose	To examine the beginning of conceptional understanding of first number words and the role of language in developing a notion of numbers
Languages	Spanish and Basque
Setting	Infant schools (including equivalent to pre-kindergarten and kindergarten, covers ages 0 – 6 years)
Number naming variations	Plural markers (Spanish) and no plural markers (Basque) associated with cardinal numbers
Participants	Female (n=70) and male (n=61) children between 2.5 and 3.5 years of age. All children had normal vision and hearing, and no reported developmental delays. Basque-speaking children were recruited from 7 infant schools in the Basque Autonomous Community (n=61), Spanish-speaking children from four infant schools in other municipalities in Spain (n=61).

Villarroel et al. (2011) investigated the early meaning and understanding of the number words 'one', 'two' and 'three' among monolingual Basque- and Spanish-speaking children, using the Give-N and How-Many tasks. The researchers aimed to understand how children acquire the concept of counting and the earliest cardinal meanings of small set-sizes associated with the dissimilar ways by which the two languages express grammatical number, namely the usage of singular/plural markers in quantities larger than one. The study included pre-school children in the early stages of developing numerical skills, aged between 2.5 and 3.5 years of age attending pre-schools in the Basque Autonomous Community and two other non-Basque-speaking regions in Spain.

While the Spanish language requires adding the suffix -(e)s to nouns and adjectives when referring to quantities larger than one, this grammatical association between cardinal quantity determiners and plural suffixes is not the general rule in the Basque language. Specifically, no plural suffixes are used in Basque references to undetermined sets of objects (for example in cases similar to the following sentences: "There are two cars", "You bought three apples", "Give me four pencils!") and optionally only the verb can be changed. Accordingly, the nouns and adjectives involved in these quantitative references are indistinguishable from their singular status.

This distinctive feature of the Basque language has its origins in the so-called indefinite grammatical case which is the general way to express quantitative aspects of undetermined sets of objects (King, 1994; Laka, 2000; Zubiri & Zubiri, 2000). Consequently, in contrast to what happens in the Spanish language, references to more than one object in Basque do not carry the compulsory use of markers to indicate their plural nature and a high proportion of communicational exchanges, which are linked to the indefinite case, lack plural references altogether, even when referring to more than one object. The research question that was addressed in this study was as follows:

- Does the exposure to a greater frequency of linguistic plural markers linked to number-words result in earlier discrimination between number 1 (singular) and the other number-words (plural) and, therefore, an earlier grasp of the first level of numerical comprehension?

The following hypotheses were articulated:

- H1: Fewer Basque-speaking children (whose language does not refer to the multiplicity of items with plural markers) will have understood the difference between one and more than one (one-knower level)
- H2: The number of subset-knowers will be correlated with age but not with sex, in line with earlier studies.

This is the second study in this scoping review which investigated the influence of semantic singular/plural markers on the understanding of the cardinality jump from one to two. This is also the second study that largely eliminated the factor of different cultural environments. Like other studies identified in this review, it did not take into account potential differences in number processing among bilingual children.

The testing consisted of Give-N and How-Many tasks. Children in this study were expected to choose a subset of a certain quantity from a set of objects when taking the Give-N test. For the How-Many task, each child was offered a set of the same objects as used before and asked to count how many objects were in the set. After having completed the Give-N-task, a child was considered to have attained a certain level of counting ability when they could show they were able to choose the specific number of objects needed for that level and that same number of objects was not selected when different subsets were requested. A maximum of one error was permitted. Children were divided into one of the three categories below based on their abilities:

- No-number-knower: incapable of correctly selecting subsets of any number.
- One-knower, two-knower, or three-knower if they could correctly select subsets of one, two, or three fruit, respectively.
- CP (Cardinality Principle)-knower: if they were able to correctly select subsets of four or more fruit.

During the How-Many-task, correct responses could be given by the children by either stating the correct number without first counting the items, counting aloud, and stopping at the last number word that corresponds to the cardinal number, or starting to count aloud, switching to silent counting, and then stating the correct cardinal number of the set at the end. Another acceptable way to respond, according to the researchers, was to simultaneously display all fingers showing the appropriate cardinal number.

The findings showed no difference between the Spanish-and Basque-speaking children in terms of their ability to use number-words for counting (How-many task) with regard to the connection between language profile and counting ability. Regarding accurately labelling the cardinality of subsets of items, comprehension of the set-size meanings of the number-words “one”, “two” and “three” varied significantly between the language groups (Figure7.5.). Children who spoke Basque were substantially more likely to lack even the most fundamental level of numerical understanding (no-number-knower) and to be unable to tell sets of ‘one’ apart from sets of other numbers (one-knower). When referring to sets of items, both groups of children understood ‘two’, ‘three” and larger numbers equally well.

Number of children that reached number-knower levels

Number-knower level	Overall	Spanish	Basque
No-number-knower	44	13	31
One-knower	58	40	18
Two-knower	17	8	9
Three-knower	8	6	2
CP-knower	4	3	1

Appendix 4: Systematic Review Summary Table

Details	Study #1	Study #2	Study #3	Study #4
Author (Year)	**Ackermans et al. (2025)**	**Akhsani et al. (2025)**	**Bang et al. (2023)**	**Byeon & Kim (2020)**
Citation	Ackermans, K., Bakker, M., van Loon, A. M., Kral, M., & Camp, G. (2025). Young learners' motivation, self-regulation and performance in personalized learning. Computers & Education, 226, 105208	Akhsani, L., Junaedi, I., & Asih, T. S. N. (2025). Exploring insights in the creative thinking process of prospective mathematics teacher students. Al-Jabar: Jurnal Pendidikan Matematika, 16(2), 581–595.	Bang, H. J., Li, L., & Flynn, K. (2023). Efficacy of an adaptive game-based math learning app to support personalized learning and improve early elementary school students' learning. Early Childhood Education Journal, 51(4), 717–732	Byeon, S., & Kim, N. (2020). Impact of Korean students' individual learning time on math performance: Differential effect of teachers' assessment competency. Asia Pacific Education Review, 21(4), 601–613
Abstract	Personalized learning, a topic that has garnered significant attention in education, is known for its potential to cater to students' unique needs and improve educational outcomes. However, most large-scale longitudinal studies on personalized learning have primarily focused on middle school students and above (age ≥ 11). This study, in contrast, delves into the uncharted territory of how personalized learning affects younger students (ages 7–12), a domain largely overlooked by large-scale studies. The objective was to understand the effect of personalized learning on young learners' academic performance, metacognitive awareness, and motivation. Multidisciplinary design teams embedded personalized learning in eight elementary schools, resulting in tailored interventions across four subjects. Effects were measured over three years among 588 students and 82 teachers and analysed using Bayesian Gaussian regression with random intercept models. Results showed significant evidence that personalized learning interventions fostered academic performance in math and spelling, that schools with explicit metacognitive skills training saw greater spelling gains, and that student ICT skills improved metacognitive awareness, intrinsic motivation, and math performance, with teachers' ICT skills supporting student metacognitive awareness. However, the theorized effect on metacognitive awareness and intrinsic motivation could not be confirmed. The study provides evidence-based recommendations for implementing personalized learning in elementary schools, particularly in math and spelling, with ICT skill development benefiting students' academic and metacognitive outcomes.	This study aims to examine the effects of differentiated instruction on primary students' mathematics achievement in the Indonesian context by comparing learning outcomes between classes taught with differentiated instruction and those taught with conventional methods.Method: This quantitative study employed a quasi-experimental, non-equivalent control group design with two intact grade 4 classes from a primary school in Indonesia. The experimental class received differentiated instruction based on students' readiness, interests, and learning profiles, while the control class received conventional teaching. Data were collected using a validated mathematics achievement test with acceptable reliability. Assumptions were checked, and independent-samples t tests compared posttest achievement between groups.Findings: Students taught with differentiated instruction obtained significantly higher posttest scores than those in the control class, indicating a positive effect on mathematics achievement.Significance: The study provides empirical evidence that differentiated instruction can improve primary mathematics achievement in Indonesian classrooms. The findings offer practical guidance for teachers and inform school leaders and policymakers seeking strategies to enhance learning outcomes	This study investigates whether using My Math Academy, which provides personalized content and adaptive embedded assessments to support existing curricula, can improve learning outcomes and engagement for kindergarten and first grade students (N = 505 treatment, 481 control). Findings indicate that students using My Math Academy made significant learning gains in math relative to children who did not. More skills mastered in My Math Academy was associated with greater learning gains on the external assessment, with the greatest impacts among students with lower levels of math knowledge and on the most difficult skills. Teachers surveyed found My Math Academy easy to use in their classrooms and recognized it as a valuable learning resource that supplemented their existing curricula to improve students' engagement, motivation, and confidence in learning math.	As the 2015 revised national curriculum is being implemented in Korea, teachers' assessment competency has become a focus for potentially enhancing students' self-directed learning. Although students' self-directed learning affects academic performance and may vary based on teachers' assessment competency and feedback practices, few studies have investigated this effect on student outcomes. Therefore, this study empirically examines the differential effect of teachers' assessment competency—especially formative assessment feedback and homework feedback—on Korean middle school students' individual learning time and mathematics achievement. Using 3-level hierarchical linear modeling (HLM) with data from 2,942 students and 426 mathematics teachers across 207 middle schools in the Gyeonggi Education Panel Study (GEPS), results show that teachers' formative assessment feedback positively influences math performance when controlling for student, teacher, and school covariates, and that the effect of individual learning time on math achievement significantly depends on homework feedback practices. The study provides new empirical evidence that teacher feedback plays a pivotal role in promoting students' self-directed learning and academic achievement
Country	Netherlands	Indonesia	USA	South Korea
Research Question	How does personalized learning affect young learners' academic performance, metacognitive awareness, and motivation?	This study aims to examine the effects of differentiated instruction on primary students' mathematics achievement in the Indonesian context by comparing learning outcomes between classes taught with differentiated instruction and those taught with conventional	Investigates whether using My Math Academy, which provides personalized content and adaptive embedded assessments to support existing curricula, can improve learning outcomes and engagement for kindergarten and first grade students	(1) Does the impact of a student's individual learning time on math performance differ due to his/her teacher's formative assessment feedback? (2) Does the impact of a student's individual learning time on math performance differ due to his/her teacher's homework(HW) feedback?

Participants	588 students completed all 3 years, ages 7 - 12	Prospective mathematics teacher students, university	954 students. 55% (n = 523) were in kindergarten, 43% (n = 406) were in first grade, and only 3% (n = 25) were in second Grade	Responses from 3,672 middle school students and 666 teachers in 314 middle schools were collected in the survey.

Details	Study #1	Study #2	Study #3	Study #4
Design and Instruments	Longitudinal, yearly measurements (Metacognitive awareness, Intrinsic motivation, and Academic performance inventory, Student ICT skills)	Quasi-experimental, non-equivalent control group design with two intact grade 4 classes from a primary school in Indonesia. The experimental class received differentiated instruction based on students' readiness, interests, and learning profiles, while the control class received conventional teaching. Data were collected using a validated mathematics achievement test with acceptable reliability.	Comparative intervention (1) a treatment condition where teachers would implement My Math Academy (21 classrooms), or (2) a control condition where teachers would implement whatever math curricula and interventions the district had in place.	Teacher and student survey design
Maths Domain	Measurements, Fractions, Ratios	General mathematics (Achievement test)	Addition, Subtraction	General mathematics (Achievement test)
Data Collection	National standardized achievement tests (math, spelling, reading, vocabulary); Junior Metacognitive Awareness Inventory; Intrinsic Motivation Inventory; student and teacher ICT-skills assessments	Numerical problem-solving tasks; written solutions; semi-structured interviews	External standardized mathematics assessment: in-app adaptive assessment and log data (skills mastered, usage); teacher surveys, observation of engagement, motivation, confidence	Panel survey data on individual learning time; math achievement scores; teacher feedback and assessment competency measures
Analysis	Bayesian multilevel Gaussian regression with random intercepts (students nested in classes/schools); MCMC estimation; Bayes factors and posterior distributions	Qualitative coding and thematic analysis aligned with stages of creative thinking (e.g., preparation, incubation, illumination, verification)	ANCOVA and regression analyses comparing learning gains; subgroup analyses by prior achievement; correlational analyses linking app usage to outcomes	Three-level Hierarchical Linear Modelling (HLM) (students nested in teachers and schools); cross-level interaction effects
Outcome	No evidence of the theorized effects of personalized learning on intrinsic motivation, no enhanced metacognitive awareness among younger students through Personalized Learning in terventions, significantly improved students' spelling and math performance. Student's (collaborative and effective) ICT skills significantly contribute to their intrinsic motivation (perceived choice), metacognitive awareness (declarative knowledge and evaluation), and academic performance (Math).	Students taught with differentiated instruction obtained significantly higher posttest scores than those in the control class, indicating a positive effect on mathematics achievement.	Findings indicate that students using My Math Academy made significant learning gains in math relative to children who did not. More skills mastered in My Math Academy was associated with greater learning gains on the external assessment, with the greatest impacts among students with lower levels of math knowledge, where there was more room for growth and on the most difficult skills. Teachers surveyed found My Math Academy easy to use in their classrooms and recognized it as a valuable learning resource that supplemented their existing curricula to improve students' engagement, motivation, and confidence in learning math.	Teacher's formative assessment feedback has a positive differential effect on math performance when controlling for student-, teacher-, and school-level covariates. In addition, the effect of individual learning time on math achievement was significantly dependent on teachers' homework feedback. The study provides newfound empirical evidence that teacher feedback plays a pivotal role in promoting a student's self-directed learning and academic achievement. Based on these findings, the implications of improving a teacher's assessment competency are discussed.

Limitations	(1) Schools followed independent development timelines for implementing interventions, leading to variations in implementation phase (design, partial, or piloting) and a lack of data on task load, frequency, or duration. (2) The three-year study period overlapped with COVID-19 lockdowns and homeschooling, which may have affected students' intrinsic motivation and metacognitive awareness and slowed intervention rollout. (3) Multilevel analyses relied on teacher-reported student motivation and metacognition, which may be biased by teachers' beliefs or unconscious stereotyping. (4) Self-report measures could have been more reliable if combined with a Think-Aloud process, but training all 721 students annually was impractical given the workload of approximately 2,000 questionnaires.	Small number of participants, limited to the numerical methods course, and was conducted over a short observation period, constrains generalizability findings	Study likely lacks detailed reporting on differential effects across subgroups (e.g., students with differing prior achievement or socio-economic status). As an educational app intervention, there is potential Hawthorne effect and limits to external validity beyond the specific app context.	While robust with large panel data, results may not fully generalize outside Korean educational contexts. Causal inference is limited due to the observational longitudinal design. The study's reliance on existing survey and administrative data may omit unmeasured variables affecting math performance (e.g., out-of-school tutoring).

Details	Study #5	Study #6	Study #7	Study #8
Author (Year)	**Febrilia et al. (2025)**	**Lai et al. (2020)**	**Lavrijsen et al. (2022)**	**Prast et al. (2018)**
Citation	Mardati, A., Machbub, A. D. Z., & Febrilia, Y. (2025). Effects of differentiated instruction on primary students' mathematics achievement: A quasi-experimental study. Al-Jabar: Jurnal Pendidikan Matematika, 16(2), 569–579.	Lai, C.-P., Zhang, W., & Chang, Y.-L. (2020). Differentiated instruction enhances sixth-grade students' mathematics self-efficacy, learning motives, and problem-solving skills. Social Behavior and Personality: An International Journal, 48(6), 1–13	Lavrijsen, J., Dockx, J., Struyf, E., & Verschueren, K. (2022). Class composition, student achievement, and the role of the learning environment. Journal of Educational Psychology, 114(3), 498–512	Prast, E. J., Van de Weijer-Bergsma, E., Kroesbergen, E. H., & Van Luit, J. E. H. (2018). Differentiated instruction in primary mathematics: Effects of teacher professional development on student achievement. Learning and Instruction, 54, 22–34.
Abstract	This study aims to examine the effects of differentiated instruction on primary students' mathematics achievement in the Indonesian context by comparing learning outcomes between classes taught with differentiated instruction and those taught with conventional methods. Method: This quantitative study employed a quasi-experimental, non-equivalent control group design with two intact Grade 4 classes from a primary school in Indonesia. The experimental class received differentiated instruction based on students' readiness, interests, and learning profiles, while the control class received conventional teaching. Data were collected using a validated mathematics achievement test with acceptable reliability. Assumptions were checked, and independent-samples t tests compared posttest achievement between groups. Findings: Students taught with differentiated instruction obtained significantly higher posttest scores than those in the control class, indicating a positive effect on mathematics achievement. Significance: The study provides empirical evidence that differentiated instruction can improve primary mathematics achievement in Indonesian classrooms. The findings offer practical guidance for teachers and inform school leaders and policymakers seeking strategies to enhance learning outcomes.	We examined the effectiveness of a differentiated instruction intervention in promoting sixth-grade students' mathematics self-efficacy (MSE), mathematics learning motives (MLM), and mathematical problem-solving skills (MPSS). The relationships among MSE, MLM, and MPSS were also assessed. A longitudinal pretest–posttest design with 25 students was used, and three instruments collected quantitative data. Findings showed that the differentiated instruction environment significantly enhanced students' MSE, MLM, and MPSS. Additionally, MSE significantly predicted both MLM and MPSS, with MLM partially mediating the effect of MSE on MPSS: stronger MSE related to higher MLM, which in turn improved MPSS	This study considers how class composition, in terms of between-student variability and the average level of achievement, is related to the academic development of students, and how these relationships can be explained by features of the class learning environment. At the start of secondary education, Flemish schools can decide autonomously how to group their students, leading to variation in class mean and class heterogeneity between classes. In a sample of 2,895 Flemish students from 158 classes, math achievement at the end of grade 8 was found to be unrelated to class heterogeneity, after accounting for previous achievement, intelligence, gender, and social background. Path analyses showed that class heterogeneity was positively associated with teachers' use of differentiated instruction to accommodate for differences between students, and that differentiated instruction was related to higher student achievement. Second, students were found to achieve better in classes with high average achievement. While this held true for all students, high achieving students seemed to benefit the most from being in a class with a high average level. Although class-average achievement was positively related to the academic orientation of the class, this did not explain the association between class mean level and achievement. These results suggest that, although it might be beneficial for the students in the high-ability groups, grouping students in distinct classes according to ability might have little overall benefit, and emphasize that teachers' responses to student diversity might be more decisive for improving student achievement than homogenizing classes in terms of ability.	This large-scale study examined the effects of a teacher professional development (PD) programme about differentiated instruction on students' mathematics achievement. Thirty primary schools (N = 5,658 students in grades 1–6) divided over three cohorts participated: Cohort 1 received the PD programme in Year 1, Cohort 2 in Year 2, and Cohort 3 was a control group. During the PD, teachers learned how to adapt mathematics education to diverse educational needs using within-class ability groups. In Year 1, the PD had a significant small positive effect on student achievement growth, with similar effect sizes for low-, average-, and high-achieving students. In Year 2, no significant effects were demonstrated. In sum, teacher PD about differentiation has the potential to promote the achievement of all students, but implementing differentiation is not straightforward, and further research is needed to identify the factors that make PD about differentiation successful.
Country	Indonesia	Taiwan	Netherlands	Netherlands
Research Question	This study aims to examine the effects of differentiated instruction on primary students' mathematics achievement in the Indonesian context by comparing learning outcomes between classes taught with differentiated instruction and those taught with conventional methods.	The effectiveness of a differentiated instruction intervention in promoting sixth graders' MSE, MLM, and MPSS.	The associations between class heterogeneity and class mean level and individual achievement, differential effects of class composition on low and high achievers, the role of the class learning environment in the association between class composition and achievement.	Whether teacher professional development (PD) about differentiation has a positive effect on student achievement in primary school mathematics.

Participants	Two full Grade 4 classes from a primary school in Indonesia.	longitudinal study, an elementary school classroom with 25 students in the sixth grade	longitudinal study, anylytic sample 2,895 students, mean age at end of Grade 8 was 14.0 years	5658 students, grades 1 - 6, Cohort 1 received the PD programme in Year 1, Cohort 2 in Year 2, and Cohort 3 was control

Details	Study #5	Study #6	Study #7	Study #8
Design and Instruments	Quasi-experimental, non-equivalent control group design. The experimental class received differentiated instruction based on students' readiness, interests, and learning profiles, while the control class received conventional teaching.	Intervention, via differentiated instruction professional development for mathematics teaching, during which the teacher designed and implemented two lessons during each semester in her classroom. The content of these lessons was related to algebra fractions and related concepts	Students completed a 5-item scale measuring the degree of differentiated instruction implemented by their math teachers and a survey to assess the perceived fit between capacities and schoolwork difficulty and to measure the degree of differentiation implemented by the math teacher; mentor teachers were surveyed on the academic orientation of their class.	Quasi-experimental, longitudinal design with a control group. Three cohorts of schools: Cohort 1: received teacher PD in Year 1 Cohort 2: received teacher PD in Year 2 Cohort 3: control group (no PD)
Maths Domain	General mathematics (Achievement test)	Algebra, Fractions	General mathematics (Achievement test)	•Numbers and number relations, covering the structure of the number line and relations between numbers •Addition and subtraction •Multiplication and division •Complex math applications, often involving multiple mathematical manipulations •Measuring (e.g., weight and length)
Data Collection	Validated mathematics achievement post-test	Mathematics self-efficacy scale; mathematics learning motivation scale; problem-solving performance test	Standardized mathematics achievement tests; student background variables; teacher-reported instructional practices	Standardized mathematics achievement tests administered longitudinally, teacher-perceived motivation
Analysis	Assumption checks followed by independent-samples t-tests comparing post-test scores	Paired-samples t-tests; mediation analysis using regression (self-efficacy → motivation → problem-solving)	Path analysis / Structural Equation Modelling (SEM) with mediation by learning-environment variables	Multilevel growth modelling (students nested in classes/schools); cohort comparisons
Outcome	Students taught with differentiated instruction obtained significantly higher posttest scores than those in the control class, indicating a positive effect on mathematics achievement.	Sixth grade student participants who received differentiated instruction in mathematics scored significantly higher on MSE, MLM, and MPSS at the end of a year-long intervention.	Math achievement at the end of grade 8 was found to be unrelated to class heterogeneity, after accounting for previous achievement, intelligence, gender, and social background. Path analyses showed that class heterogeneity was positively associated with teachers' use of differentiated instruction to accommodate for differences between students, and that differentiated instruction was related to higher student achievement. Second, students were found to achieve better in classes with high average achievement. While this held true for all students, high achieving students seemed to benefit the most from being in a class with a high average level. Although class-average achievement was positively related to the academic orientation of the class, this did not explain the association between class mean level and achievement. Although it might be beneficial for the students in the high ability groups, grouping students in distinct classes according to ability might have little overall benefit, and emphasize that teachers' responses to student diversity might be more decisive for improving student achievement than homogenizing classes in terms of ability.	In Year 1, the PD had a significant small positive effect on student achievement growth. The effect size was similar for low-achieving, average-achieving and high achieving students. In Year 2, no significant effects were demonstrated. In sum, teacher PD about differentiation has the potential to promote the achievement of all students. However, implementing differentiation is not straightforward and future research is necessary to unravel which factors make PD about differentiation succeed.

Limitations	As a quasi-experimental study, assignment to intervention and control conditions wasn't fully randomized, which can introduce selection effects. This design lacks the high internal validity of a randomized controlled trial, and potential confounds such as teacher enthusiasm, classroom routines, or unmeasured prior achievement might have affected outcomes. In addition, the results are limited to a specific geographic primary school population.	Small sample size reduces statistical power and limits generalizability. The pretest–posttest design lacks a control group, so maturation or test-practice effects can't be ruled out.	Class composition effects may be confounded with unmeasured classroom practices or school policy differences. Although the study controls for prior achievement and background, context-specific factors in Flemish education may limit broader generalization.	Small effect sizes in Year 1 only; no significant Year 2 effects; implementation challenges not fully understood; need for further research on factors that make PD effective.

Details	Study #G	Study #10	Study #11	Study #12
Author (Year)	**Rodríguez-Martínez et al. (2023)**	**Zerai et al. (2023)**	**Zhang et al. (2017)**	**Ziegler et al. (2021)**
Citation	Rodríguez-Martínez, J. A., González-Calero, J. A., del Olmo-Muñoz, J., Arnau, D., & Tirado-Olivares, S. (2023). Building personalised homework from a learning analytics based formative assessment: Effect on fifth-grade students' understanding of fractions. British Journal of Educational Technology, 54(1), 76–97.	Zerai, D., Eskelä-Haapanen, S., Posti-Ahokas, H., & Vehkakoski, T. (2023). The use of question modification strategies to differentiate instruction in Eritrean mathematics and science classrooms. Education Sciences, 13(3), 284	Zhang, D., Ding, Y., Lee, S., & Chen, J. (2017). Strategic development of multiplication problem solving: Patterns of students' strategy choices. The Journal of Educational Research, 110(2), 159–170	Ziegler, E., Edelsbrunner, P. A., & Stern, E. (2021). The benefit of combining teacher-direction with contrasted presentation of algebra principles. European Journal of Psychology of Education, 36(1), 187–218
Abstract	This study analyses the potential of a learning analytics (LA)-based formative assessment to construct personalised teaching sequences in mathematics for 5th-grade primary school students. A total of 127 students from Spanish public schools participated in the quasi-experimental study conducted over the course of six sessions, in which both control and experimental groups followed a teaching sequence of mathematical problems about fractions. In each session, both groups used audience response systems to record their responses. After each session, the control group received generic homework on fractions, while the experimental group received personalised homework based on errors detected through the LA-based formative assessment. After the intervention, results indicated a higher level of understanding of fractions in the experimental group compared to the control group. Additionally, instruction using audience response systems had a positive effect on motivational dimensions compared to regular mathematics classes.	This qualitative study aimed at examining the question modification strategies Eritrean elementary and middle school teachers used to differentiate their instruction and meet the diversity in the classroom as well as the functions these strategies served in classroom interactions. The research data consisted of videotaped recordings (N = 11 videotaped lessons) of classroom interactions in eight mathematics and science classrooms, which were analysed through interaction analysis. The findings showed that Eritrean teachers utilised the following five question modification strategies either independently or in combination: repetition; rephrasing; clarification; decomposition; and code-switching. Although repetition was the most commonly used strategy, it was not found to help teachers to differentiate their instruction. Likewise, the utilisation of rephrasing was dependent on how effectively teachers captured students' misunderstandings and modified their questions accordingly. Instead, clarification, decomposition, and code-switching were found to be the most highly developed question modification strategies from the viewpoint of differentiation. It was concluded that the question modification strategies were dominant and workable elements of classroom interactions in teacher-led and poorly-resourced large classrooms, such as those in Eritrea	Low mathematics achievement is a persistent problem in the United States, and multiplication is a fundamental area in which many students manifest learning difficulties. This study examined the strategic developmental levels of multiplication problem solving among 121 elementary school students in Grades 3 through 5. A latent class analysis modeling was used to identify three valid groups representing different patterns of strategy choices for each of three types of multiplication problems. Findings indicated intra-group variability for problem-solving accuracy, for frequency of using different strategies, and for accuracy of executing direct retrieval/algorithm (DR/AG) strategies. Students demonstrated relative consistency in their strategy choices for solving the three problem types. Students who used DR/AG strategies most frequently showed the highest problem-solving accuracy and the highest accuracy of executing the DR/AG strategies. Students who most frequently relied on incorrect operations or who indicated they did not know how to solve problems demonstrated the lowest problem-solving accuracy among the three groups; the number of students in this group increased with problem difficulty levels. Implications are discussed in terms of identifying students' strategic developmental levels and providing differentiated instruction based on the identified levels	Teacher-directed and self-directed learning have been compared across various contexts, with mixed benefits depending on setting and presentation of material. This study re-analysed data from two experimental studies investigating the effects of contrasted versus sequential presentation of algebraic addition and multiplication principles in sixth-grade classrooms. Both groups were presented the same structured materials differing only in whether the principles were explained by the teacher (teacher-direction) or inferred by the students (self-direction). The results showed short-and medium-term advantages of combining teacher-direction with contrasted presentation of algebra principles. Moreover, examination of aptitude-treatment interactions revealed that students with lower reasoning abilities particularly benefited from teacher-direction in the contrasted condition. Based on these findings, the authors discuss circumstances under which teacher-directed instruction reveals its advantages
Country	Spain	Eritrea	USA	Switzerland
Research Question	RQ1: Does personalised homework from an LA-based formative assessment have a significant effect on fifth-grade students' conceptualization of fractions in comparison with non-personalised homework? RQ2: Do ARS-based teaching sequences have a significant effect on fifth-grade students' motivational variables compared to non-ARS-based teaching sequences?	(1) What kinds of question modification strategies do Eritrean teachers use in mathematics and science classrooms to differentiate their instruction? (2) What functions do various question modification strategies serve in differentiating instruction?	(a) Are there varying classes among elementary students that show different patterns of strategy choices and, if yes, what are the major characteristics of each class? (b) How valid is the classification? (c) How consistent is the classification of these elementary students for solving different types of multiplication problems?	Whether learning is more efficient when the teacher explicitly outlines principles, rules, solutions and concepts, or when students infer and discover them on their own.
Participants	127 fifth-grade students from five classes in two public schools in Spain.	The data included five elementary school classrooms (grades 4 and 5) and three middle school classrooms (grades 6)	121 students, 52 third-grade students, 34 fourth-grade students, and 35 fifth-grade students.	250 students, Sixth graders with no prior algebra instruction

Design and Instruments	Quasi-experimental study with four mathematical sessions (see Figure 1). The mathematical sessions were identical for both groups and ARS were used to work on different mathematical problems related to fractions in both groups.	Interaction analysis of videotaped recordings (11 lessons) of classroom interactions in eight mathematics and science classrooms.	Three test probes of multiplication calculation problems were administered to the participants, including a 1 £1 digit multiplication test, a 1 £ 2 digit multiplication test, and a 2 £ 2 digit multiplication test, with 10 problems on each test.	A total of 14 classes participated: 9 classes received self-directed instruction, and 5 classes received teacher-directed instruction.

Details	Study #G	Study #10	Study #11	Study #12
Maths Domain	Fractions	Classroom interactions on maths topics	Latent Class Analysis (LCA)	Algebraic operations
Data Collection	Audience Response System (clicker) data; fraction-understanding tests; motivation questionnaires Reduced Instructional Materials Motivation Survey (RIMMS) (attention, relevance, confidence, satisfaction)	Qualitative video observation and interaction analysis of classroom discourse (11 lessons); lesson transcripts	Written multiplication problem-solving tasks; coded strategy-use data	Pre-, post-, and delayed post-tests on algebraic principles
Analysis	ANCOVA and regression analyses comparing experimental and control groups	Interaction analysis / qualitative content analysis identifying and categorizing question-modification strategies	Latent Class Analysis (LCA) to identify strategy profiles; descriptive and comparative analyses of accuracy	ANOVA and aptitude–treatment interaction analyses examining short- and medium-term effects
Outcome	the group provided with personalised homework from an LA-based formative assessment on domain-specific didactic knowledge of fractions attained a higher level of understanding related to specific kinds of fractions knowledge than the students given non-personalised homework on fraction concepts. In addition, our results confirm the need to address different activities when working with fractions, at elementary levels, adapted to the individual needs of each student	Teachers utilised the five question modification strategies either independently or in combination: repetition; rephrasing; clarification; decomposition; and code-switching. Although repetition was the most commonly used strategy, it was not found to help teachers to differentiate their instruction. Likewise, the utilisation of rephrasing was dependent on how effectively teachers captured students' misunderstandings and modified their questions accordingly. Instead, clarification, decomposition, and code-switching were found to be the most highly developed question modification strategies from the viewpoint of differentiation. It was concluded that the question modification strategies were dominant and workable elements of classroom interactions in teacher-led and poorly-resourced large classrooms, such as those in Eritrea.	Findings indicated intragroup variability for problem-solving accuracy, for frequency of using different strategies, and for accuracy of executing direct retrieval/algorithm (DR/AG) strategies. Students demonstrated relative consistency in their strategy choices for solving the three problem types. Students who used DR/AG strategies most frequently showed the highest problem-solving accuracy and the highest accuracy of executing the DR/AG strategies. Students who most frequently relied on incorrect operations or who indicated they did not know how to solve problems demonstrated the lowest problem-solving accuracy among the three groups; the number of students in this group increased with problem difficulty levels. Implications are discussed in terms of identifying students' strategic developmental levels and providing differentiated instruction based on the identified levels.	The contrast teacher-directed condition resulted in the highest performance, and the sequential self-directed condition resulted in the lowest performance. The sequential teacher-directed condition and contrast self-directed condition resulted in similar outcomes, indicating similar added values of teacher-directed compared to self-direct instruction and of contrasted compared to sequential order.
Limitations	Intervention effects vary over time and may not be sustained without ongoing support. PD implementation fidelity and teacher uptake was variable, potentially reducing effect size and complicating interpretation of causal effects across cohorts.	Learning analytics homework personalization studies often face issues of implementation fidelity, data quality limitations (e.g., only capturing digital homework behavior), and short intervention duration, limiting long-term effect claims.	Strategy choice studies typically rely on cross-sectional designs, limiting causal inference. Results may depend on measurement methods for strategy identification and may not capture longitudinal development.	Combined often experimental designs with teacher-direction studies can have context-specific teacher effects and classroom variability that limit generalizability. Small or non-random samples can also reduce external validity.