Doctoral theses at NTNU, 2024:4

Gunnhild Saksvik-Raanes

Digital assessment of early number sense

The design and validation of an assessment tool

Doctoral thesis

NTNU
Norwegian University of Science and Technology
Thesis for the Degree of
Philosophiae Doctor
Faculty of Social and Educational Sciences
Department of Teacher Education

NTNU
Norwegian University of
Science and Technology

Gunnhild Saksvik-Raanes

Digital assessment of early number sense

The design and validation of an assessment tool

Thesis for the Degree of Philosophiae Doctor

Trondheim, January 2024

Norwegian University of Science and Technology
Faculty of Social and Educational Sciences
Department of Teacher Education

NTNU
Norwegian University of
Science and Technology

NTNU
Norwegian University of Science and Technology

Thesis for the Degree of Philosophiae Doctor

Faculty of Social and Educational Sciences
Department of Teacher Education

ISBN 978-82-326-7608-8 (printed ver.)
ISBN 978-82-326-7607-1 (electronic ver.)
ISSN 1503-8181 (printed ver.)
ISSN 2703-8084 (online ver.)

Doctoral theses at NTNU, 2024:4

Printed by NTNU Grafisk senter

Preface

The other day, my youngest son asked me, "Mom, do we ever get too old to learn new things?” As so many other times, a question from a 5-year-old sets off the most interesting conversations. I have been so lucky to start my PhD from a problem I encountered in my teaching practice, working with the fascinating mathematical ideas of 5 and 6-year-old children before developing an assessment tool to try to describe these ideas while learning so many new things along the way. Developing an assessment is not a straightforward process, nor is writing about it. Thankfully I have not done this alone, and consequently, many deserve my utmost gratitude for helping me: Thank you to my main supervisor; Associate Professor Trygve Solstad, for encouraging me to start this journey and for all your hard work, support, and patience. Thank you to my co-supervisor; Professor Jeremy Hodgen, for sharing your theoretical insights and perspectives on assessment and for helping me to connect the story of my thesis. Thank you to co-supervisor; Associate Professor Yvonne Grimeland, for participating in the development of ENSA, for your mathematical expertise and for the, at times, sorely needed moral support.

Participation in courses at NTNU and other institutions has expanded my knowledge while challenging my perspectives. I am grateful for everything I have learned about measuring mathematical competence at the University of Oslo and especially from Professor Guri Nortvedt, her participation in expert review groups, discussions and encouragement throughout the process. Thank you to the University of Agder for providing me with theoretical and methodological insights into research in mathematics education and especially to Professor Simon Goodchild for helping me find the connections between the process and product of learning. Additionally, I am grateful for everything I have learned about Rasch measurement at the University of Illinois at Chicago and from Professor Everett Smith.

Many people have contributed in their unique way to realising this project and deserve my gratitude; my research group at NTNU and particularly Tore Alexander Forbregd for invaluable contributions to the development of ENSA; Eivind Kaspersen for support on test development, data analysis, interpretations and for sharing your thoughts about “Enjoying your time as a PhD student”; Bjørn Evjen for programming and scoring; the children who shared their thoughts while carrying out the assessment and the teachers for their reflections; the six master students for daring to write their master thesis on unfamiliar grounds; the reviewers who has commented on my drafts – helping me to become a better writer; my fellow PhD-students for making me feel less alone in a world of self-supposed imposters striving for perfection until realising that perfect does not really exist; and finally to Astrid Junker for contributing to the development process, data collection, and so many other things - you know I could never have done this without you.

Finally, a thank you to those who matter the most. My sister and “secret supervisor”, Professor Ingvild Saksvik-Lehouillier, for allowing me to ask every stupid question. My little brother Simen Berg Saksvik for encouraging me by finishing his PhD before me. Now I am no longer the only one of us without a PhD. Thank you to my father and Professor Per Øystein Saksvik for reminding me that there are things in life that are more important than one’s career. My mother, Siri Berg, for encouraging me to be ambitious and optimistic about what a mother of three can manage - now you are the only one in the family without a PhD. This thesis is dedicated to my family. To Martin for keeping up with me and understanding when I did not even understand myself, and to Viktor, Iris, and Vetle for being my world. Now the work is done, and I am all yours again.

Abstract

Children enter school with several valuable mathematical experiences, but many teachers lack the methods to formatively assess young children's mathematical knowledge. Consequently, children may not be met with sufficient challenges to facilitate their further mathematical development. Moreover, current assessment tools used for the youngest children are either too time-demanding to carry out, lack focus on specific mathematical domains or do not consider various aspects of assessment validity. In this PhD thesis, I investigate how technology can enhance the formative assessment of children's early number sense by developing a digital assessment tool—early number sense assessment (ENSA)—focusing specifically on early number sense and assessment validity. The development and further investigation of ENSA in use reveals that technology can supplement the assessment of number sense in a time-efficient way without being dependent on reading and writing skills or the competence of the individual instructor. ENSA also describes previously unknown aspects of Norwegian children's early mathematical knowledge. Furthermore, the thesis shows how interactive items have the potential to describe more of the process, leading to an answer to mathematical problems while displaying evidence of assessment validity. Additionally, the use of technology brings some specific design elements that can affect children's results and strategies when finding answers to mathematical problems. Finally, teachers' interpretations of children's assessment results situated in a wider social practice are limited and shaped by external factors that ultimately affect the formative value of assessments for teachers.

Table of Contents

1 Introduction

1.1 Young children's mathematical knowledge when they start school

To facilitate differentiated teaching, basing teaching and learning in the first grade on children's previous knowledge, teachers need to describe children's number sense when they start school. We know that children come to school with valuable mathematical experiences and knowledge, but these experiences are not always reflected in the activities with which children meet in school (Bisanz et al., 2005; Clarke et al., 2006; Litkowski et al., 2020). There is also large variability in children's number-related knowledge before they start school (Saksvik-Raanes et al., 2023). These differences may affect children's further learning opportunities (Duncan et al., 2007) and are related to equity (Zhu, 2018). Moreover, misalignment between children's knowledge and the activities they are introduced to in school may negatively affect their mathematical development (Bodovski & Farkas, 2007). One way to meet these challenges is to learn more about what children know using assessment tools (Devlin et al., 2022; Raudenbush et al., 2020). However, many teachers struggle with finding the time and tools to effectively evaluate the mathematical knowledge of 5- and 6-year-old children.

1.2 Formative assessment tools for describing children's knowledge

Numerous assessment tools are available to evaluate aspects of children's mathematical understanding (Clements et al., 2008; Ginsburg & Baroody, 2003; Ginsburg & Pappas, 2016; Jordan et al., 2010; van de Rijt et al., 1999). How these different tools are developed and used may affect the teaching and learning of mathematics in schools. One weakness of many available assessment tools is that they do not provide specific results about what children already know or their further learning potential. Additionally, many tests do not provide information about the process leading up to an answer. Moreover, current assessments are largely affected by children's reading and writing skills or by time-consuming assessment methods, such as individual interviews.

1.3 The potential of technology for assessing children's number sense

New technological developments provide the potential to create tools that can shed light on aspects of children's understanding and consequently influence teaching practices (Ginsburg, 2016). Digital assessment offers several advantages over written assessment, especially in assessing children's mathematical knowledge (Drijvers, 2018). The use of technology brings possibilities to make assessment items more accessible for children, to look deeper into children's mathematical knowledge and to describe further aspects of children's solution process than what is possible in a static pen-and-paper format (van den Heuvel-Panhuizen et al., 2011).

Making assessment items more accessible for young children implies not being dependent on the individual child's reading or writing skills and enabling parallel delivery without the need to pace the children (Saksvik-Raanes et al., 2023). Additionally, technology enables the adaptation of complex items previously available in the assessment format using more time-demanding methods, such as individual interviews (Saksvik-Raanes & Solstad, 2021). Motivation is also central in making assessment items more accessible to children. Recent research has described how digital tools might make learning and assessment enjoyable for children (Ginsburg et al., 2019; ten Braak & Størksen, 2021).

Looking deeper into children's mathematical knowledge and describing further aspects of their solution process might imply including interactive items (Saksvik-Raanes &

Solstad, 2023) and using log data on children's behaviour to provide insights into the processes leading to answers (Bostic et al., 2019; Fischer et al., 2020; Olsher et al., 2016). For instance, including screencasts in formative assessments can provide teachers with more information regarding children's mathematical understanding than the information provided by pen-and-paper assessments (Soto & Ambrose, 2016).

In this context, there is a need to consider how technologically initiated changes in the response format, interactivity and various representations might affect how assessment tools are utilised to describe children's understanding and further learning potential. At the same time, there is a gap in the recent research literature regarding how technology affects assessment validity. Do technologically enhanced assessment tools assess the same type of understanding as traditional assessment tools, and can these tools tell us more about children's number sense without compromising assessment validity?

This PhD thesis originates from teacher practice and the need to assess children's number experiences at the starting point of their formal education. As a teacher in the Norwegian primary school, I found it challenging to find assessment tools that could describe the individual differences in the children's number sense in a time-efficient way that simultaneously displayed evidence of validity. Therefore, in my PhD project, I have developed a specific digital resource to evaluate children's number sense and investigate how this resource can contribute to the formative assessment of first-grade children's number sense in a time-efficient way, with a focus on several aspects of validity. I have designed the early number sense assessment (ENSA) tool to help teachers describe children's number sense.

ENSA was developed using an item response modelling approach, operationalising number sense by applying the foundational number sense (FoNS) model, enabling the description of children's number sense through eight different categories. Furthermore, several validity aspects of the tool were investigated by following the descriptions of an instrument development process oriented towards identifying validity evidence (Wolfe & Smith, 2007), specifically focusing on instructional validity (Pellegrino et al., 2016). With an emphasis on formative assessment practices, I investigate how the results of this assessment process can be used to promote further learning. In this way, ENSA enables us to address some important questions related to the use of technology in assessments and the central aspects of assessment validity.

Taking a mixed methods approach in a design-based research methodology, I have investigated the validity evidence of the assessment tool by applying validation activities in a Rasch measurement context. The final tool consisted of 76 items presented to the children using touchscreen tablets. A total of 77 five-year-old children in their final year of Norwegian early childhood education and 368 first-grade children at the start of Norwegian primary school participated in the study. The children's responses were analysed using the dichotomous Rasch model to evaluate the quality of the assessment tool. Five teachers were interviewed while being presented with children's responses to ENSA to explore further validity aspects related to consequential validity and how teachers interpret and use assessment results.

1.4 Research questions

The following main research question guides the study:

How can a specific digital resource (ENSA) contribute to the formative assessment of children's number sense at the beginning of first grade?

This question is addressed through the five papers in the thesis with the following research questions:

1) How do digital tasks support aspects of validity in the assessment of first graders' number sense?
2) Which design elements influence the level of difficulty of digital arithmetic assessment items, and how do the design elements influence the strategies that first-grade children use to solve these items?
3) What can a digital assessment tool tell us about five-year-old children's early number sense?
4)
    a. What is the evidence for the validity of ENSA – a digital tool for assessing early number sense?
    b. Can the interactive items be considered part of the same construct as the regular multiple-choice items?
    c. What is the added value of including interactive items in the assessment?
5) What factors influence teachers' interpretations when reflecting on the assessment of children's number sense?

1.5 Structure of the thesis

This thesis is structured around five papers that shed light on the relevant stages of the development and validation of ENSA. I start framing the five papers by describing the current literature on children's number sense learning, formative assessment and assessment tools, in addition to teachers' assessment knowledge. Through this literature review, the conceptual frameworks that lay the foundation for the development and investigation of ENSA are identified. Furthermore, I present the theoretical frameworks related to learning about numbers in a social context, assessment of previous learning and assessment validity. Next, I describe the methodological framing and methods used to investigate how ENSA can contribute to the formative assessment of children's number sense. In the methods chapter, I go into more detail on the development process of ENSA and the applied methods for investigating how the tool can contribute to the formative assessment of children's number sense. In Chapter 6, I provide a detailed summary of the five papers in the thesis. In the final chapter of the thesis, I summarise the findings and describe how each manuscript contributes to answering the presented research questions. I also discuss how the thesis provides conceptual, methodological and empirical contributions to mathematics education research before I look at the limitations and implications of my findings and further work.

2 Literature review

I need to place my study within the current research literature to investigate how a specific digital resource (ENSA) can contribute to the formative assessment of children's number sense. In this chapter, I look at the different elements of my research question to guide the literature review. I start at the final part of my main research question: *children's number sense at the beginning of first grade*. Here, I introduce different perspectives on the number sense concept and related notions to define the knowledge described in the assessment process. I will also look into the current assessment practices for describing children's number sense using formative assessment and related tools. Finally, to address the part of my research question related to the formative assessment tool's contributions, I describe research related to the knowledge that teachers use in their formative assessment practices to facilitate further learning.

2.1 Number sense: A collection of basic skills or a deep understanding?

Various descriptions exist of what constitutes children's early mathematical knowledge. Some refer to this knowledge as number sense; others might call it numeracy. Informal mathematical knowledge is also used to describe children's first attempts to quantify the world around them. Many research fields have been interested in describing children's early number knowledge from different perspectives, resulting in various concepts and definitions. Because of the variations in defining the term number sense, Jordan et al. (2009) has previously used the term number competence to emphasise that these are skills that can be learned.

Number sense can be described as a body of knowledge and a way of thinking (Sowder & Schappelle, 1989). On one level, number sense is described as an ability to compare and judge the reasonableness of calculations, do mental computations and estimate. On another level, number sense is a way of thinking or a "well-organized conceptual network that enables a person to relate number and operation properties" that is not easy to teach or measure (Sowder & Schappelle, 1989, p. 4). This relates to a description of number sense as the general understanding and use of numbers and operations in flexible ways to develop strategies, communicate, process, and interpret information (McIntosh et al., 1992). Reviewing the current literature, Berch (2005) revealed that number sense was a complex construct ranging from descriptions of skills or abilities (e.g. to estimate or decompose numbers) to conceptual structures or processes. Within the latter category of descriptions, number sense was viewed as a deep understanding of mathematical principles and relations and an appreciation of mathematical consistencies and regularities.

In the following subchapters, I will extend some of these descriptions of number sense as a skill and a deep understanding. Number sense skills and the related components will be connected to notions of numeracy (Aunio & Räsänen, 2016), number competence (Devlin et al., 2022) and informal mathematical knowledge (Milburn et al., 2019) to give a broader overview of current research on children's early number knowledge. Furthermore, I will connect the descriptions of number sense as a conceptual network (Sowder & Shappelle, 1989) and understanding (McIntosh et al., 1992) to related frameworks on mathematical knowledge. Finally, I will introduce three related perspectives on number sense and how these relate to previous perspectives on the concept. Because of my focus on children's number sense at the start of Norwegian primary school, all of the introduced research will mainly relate to children's early mathematical knowledge between the ages of 5 and 7.

2.1.1 Number sense skills and related components

The skills that are included in descriptions of number sense often involve five specific counting principles (Gelman, 1978): 1) The one-one principle includes rhythmic coordination of partitioning and marking objects and keeping track of counted objects as well as connecting one component (e.g. number word) to another component (e.g. specific object). 2) The stable-order principle includes knowing the count list and the stable as well as the repeatable order of numbers. Ordinality is also related to the number sense concept, which includes knowledge of the count list. 3) The cardinal principle relates to the knowledge of "how many", connected to the fact that the final tag in the tagging process has a specific significance, representing the total amount in the set. 4) The abstraction principle implies knowing that the previously described principles can be applied to anything to be counted. 5) The order-irrelevance principle states that the order of tagging objects or counting is irrelevant. These five principles, elements or variations are represented within different descriptions of children's number sense and numeracy.

Recent research describing the elements of number sense includes three interconnected strands: number (including counting), number relations and number operations. Devlin et al. (2022) refer to number, number relations and number operations as subdomains of early number competence. While there are variations in the descriptions of how these strands or subdomains are elements of number sense or number competence, the three strands reinforce each other, predict further achievement, and individual differences (Jordan et al., 2022). The number sense strands may be connected to a previously introduced framework for considering number sense focusing on numbers, operations and computational settings (McIntosh et al., 1992). Abilities included in various descriptions of number sense can also be seen in descriptions of numerical skills. Aunio and Räsänen (2016) described a model identifying central numerical skills within four groups that are fundamental in developing children's mathematical understanding between five and eight years. The skills identified from widely used standardised mathematical tests are described in the model as 1) having symbolic and non-symbolic number sense, 2) understanding mathematical relations, 3) having counting skills and 4) possessing basic arithmetic skills. In this model, counting skills are included in the descriptions of numeracy. This model is similar to the interconnected strands of number sense introduced by Jordan et al. (2022). However, similar to many other descriptions of number sense, the model of Jordan et al. (2022) emphasises interconnectedness and flexibility within the strands of number sense.

Based on a constant comparison analytical approach to the literature on number sense development related to grade one children, Andrews and Sayers (2015) identified eight flexible and relational categories of FoNS: 1) number identification, 2) systematic counting, 3) number-quantity relationships, 4) quantity discrimination, 5) representing numbers, 6) estimation, 7) simple arithmetic competence and 8) awareness of number patterns. These categories are related to one of three aspects of number sense identified by Andrews and Sayers (2015). Within one of these perspectives, the eight categories of FoNS are viewed as relational, describing the different aspects of number sense within one component. These perspectives will be described further in Chapter 2.1.3.

Others emphasise the componential nature of children's number-related knowledge (Dowker, 2005, 2016;Jordan et al., 2022). Dowker (2005) highlighted nine components of 6- and 7-year-old children's early numeracy, as emphasised by previous research and teachers: 1) counting procedures, 2) counting principles, 3) written arithmetical symbols, 4) place value in arithmetic, 5) understanding and solving word problems, 6) translation between verbal and numerical format, 7) calculation with derived fact strategies, 8) arithmetical estimation and 9) memory of number facts (p. 248). Jordan et al. (2022), on the other hand, drew on studies describing the previously introduced three-factor model.

In a broader context, informal mathematical knowledge has been described as a complex structure of constructs (Milburn et al., 2019, p. 493). Informal mathematical knowledge often refers to the mathematical knowledge children acquire in their everyday lives without using written symbols and algorithms (Baroody et al., 1984). Milburn et al. (2019) explained the multidimensionality of informal mathematical knowledge using a four-factor model that contains numbers and operations, measurement, geometry and patterning. One of these four factors, the number and operations factor, includes three first-order factors: numbering, operations and relations. The descriptions of Milburn et al. (2019) also relate to previous studies investigating preschool children's informal numeracy skills. Purpura and Lonigan (2013) assessed preschool children on an early numeracy test to identify a model that could represent the structure and relations of children's informal numeracy skills. They found a three-factor model comprising distinct but related factors: numbering, relations and arithmetic operations (Purpura & Lonigan, 2013). A contrast to these findings is the investigation of Research-Based Early Maths Assessment development using the Rasch model. Clements et al. (2008) found early mathematical achievement to be a unidimensional construct.

While previous research highlights the various components of number sense, numeracy and children's informal mathematical knowledge, there are differences in the descriptions of these components and how they relate. At the same time, many of the described studies point out the importance of further investigating the structures of children's early mathematical knowledge and how these components develop over time and affect children's learning from a longitudinal perspective. In their investigations of children's informal numeracy skills, Purpura and Lonigan's (2013) approach enabled them to compare different models when describing the structure of the construct. Such an approach could facilitate further investigation of children's early mathematical development. In this way, we can secure better descriptions of the number sense construct and how the various components affect children's development to facilitate further learning. At the same time, many of the components of the various definitions and investigations of children's early mathematical understanding are related to wider descriptions of mathematical knowledge. I will look more deeply into these descriptions in the next chapter.

2.1.2 Number sense and mathematical understanding

A further level in many introduced definitions involves descriptions of number sense as a deep understanding, a conceptual network or structure, including flexibility and fluency with numbers. These descriptions relate to more general descriptions of mathematical understanding. Many terminologies exist when it comes to analysing what it means to master mathematics and explaining how knowing or doing mathematics is more than just knowing the procedures and applying the skills (Niss et al., 2016). *How* children find answers to mathematical problems relates to children's mathematical understanding (Skemp, 1976), procedural and conceptual knowledge (Hiebert & Lefevre, 1986), mathematical proficiency (Kilpatric et al., 2001) and further strategy adaptivity (Verschaffel et al., 2009). These concepts are central in defining what applying number sense to solve mathematical problems implies. I will describe how I define 1) mathematical understanding, 2) procedural and conceptual knowledge, 3) mathematical proficiency and 4) strategic adaptivity before I, in subchapter 2.1.3, connect the four perspectives to the number sense concept and three related perspectives on number sense.

There are different interpretations of the words *understanding* and *knowledge*. Although knowledge and understanding are defined similarly as someone's knowledge or understanding regarding a specific phenomenon and are used as synonyms, they might also be used differently. In this context, I describe the word *understanding* within Skemp's (1976) reflections on instrumental and relational understandings. I also look at *knowledge* as procedural and conceptual, as described by Hiebert and Lefevre (1984). However, the concepts of knowledge and understanding are used interchangeably in the rest of this thesis.

More than four decades ago, Skemp (1976) introduced two types of faux amis in a mathematical context: the word *understanding* and the word *mathematics* itself. According to Skemp (1976), the different meanings of the word understanding were, at this time, the root of difficulties in mathematical education. The two meanings of understanding were described as relational and instrumental. A relational understanding was defined as "knowing what to do and why". In contrast, an instrumental understanding was described as "rules without meaning" (Skemp 1976, p. 20). An even more serious faux amis described by Skemp (1976) was mathematics, where two different types of subjects were taught under the name mathematics. What constitutes mathematics was, in this sense, not just the content of the mathematical subject but the way of knowing about it. Skemp (1976) described relational mathematics as a mental map and instrumental mathematics as successive stages without relationships with the final goal. The relational understanding was related to building structures that could produce unlimited plans to reach the finishing point. Therefore, the two kinds of understanding could be considered two different kinds of mathematics. Expectations from exams and tests and the difficulty of assessment were described as situational factors contributing to the difficulty of developing a relational mathematical understanding. In the assessment context, we would be interested in knowing more about what type of understanding children apply to solve mathematical problems and how situational factors in the assessment process might affect children's mathematical understanding.

In mathematics education, children's *knowledge* is often described as procedural or conceptual. Procedural knowledge is frequently referred to as the formal mathematical language and algorithms of mathematics that can be learned with or without meaning. In contrast, conceptual knowledge is described as a connected network of knowledge, allowing for more abstract levels of thinking (Hiebert & Lefevre, 1986). In this sense, conceptual knowledge is connected to deeper knowledge connected to real mathematics and creativity than what is said about procedural knowledge. Thus, one might relate conceptual knowledge to previous descriptions of relational understanding. At the same time, procedural knowledge may also be seen as flexible in adapting procedures related to conceptual knowledge (Star, 2005). Consequently, procedural knowledge is also important for developing deeper mathematical knowledge. Previous studies have found that comparing solution processes affects both children's procedural and conceptual knowledge in a positive way (Rittle-Johnson & Star, 2007). A focus on strategies and means for comparing these may thus contribute to children's procedural and conceptual knowledge development.

The meaning behind the concept of *strategy* depends on the applied theoretical framework, which I will return to in Chapter 3. From a constructivist point of view, strategy encloses the means an individual uses to solve a mathematical problem. From a broader perspective, strategic adaptivity implies consciously or unconsciously choosing a solution strategy adapted to the mathematical problem and sociocultural context (Verschaffel et al., 2009). One might also look at the emergence of strategies as more than a choice where the flexibility lies in how the individual interacts with the problem and experiences the possibilities of the numbers rather than applying a specific solution path (Threlfall, 2002). In this sense, the interaction with the problem and its mathematical meaning is in focus (Proulx, 2013), with various ways children may experience the numbers in the situation (Björklund & Kempe, 2022). In the context of this thesis, I focus mostly on the children's strategies from an individual constructivist perspective when describing how the individual children solve the items in ENSA. However, I also look at the children's strategies from a broader perspective when reflecting on how their interactions with the interactive environment affect their strategies.

The various types of knowledge, understanding and choice of strategies are related to a person's competence. Competence can be viewed from an analytical or holistic viewpoint (Blömeke et al., 2015). Defining the term holistically implies including broader terms of the construct, considering its more complex characteristics. With the notation competency, an analytical viewpoint means dividing the skills into different constitutes involving feelings or skills. Both perspectives enhance the importance of looking at typical situations where competence is needed and that this is recognised in the framework (Blömeke et al., 2015). Because of this complexity, it is difficult to find a precise definition of the term competence. Koeppen et al. (2008) conceptualised competence as complex and context-specific trainable ability constructs that are "closely related to real life" (p. 61). I will lean towards this definition of competence, highlighting the aspects of real-life stations. In addition, Blömeke et al. (2015) pointed out that a definition of competence should include both complex cognitive abilities and affective dispositions guided by an individual's will. Therefore, the applied definition of competence must also emphasise affective dispositions and dispositions related to the individual.

Various competencies are needed to enable children to perform mathematics (Kilpatrick, 2014). I have restricted my understanding of mathematical competence to mathematical proficiency, as Kilpatrick et al. (2001) described, capturing what it means to learn mathematics successfully. With five interwoven and interdependent strands, mathematical proficiency reflects a person's conceptual understanding, procedural fluency, strategic competence, adaptive reasoning and productive disposition (Kilpatrick et al., 2001). Within mathematical proficiency, conceptual understanding can be connected to the described perspectives on relational understanding and conceptual knowledge – referring to grasping mathematical ideas and connecting them to previous knowledge. Procedural fluency might be related to descriptions of procedural knowledge and skills in performing flexible and efficient procedures adapted to mathematical problems. Furthermore, strategic competence relates to solving mathematical problems through various formulations and representations. Adaptive reasoning includes reasoning and justifying mathematical processes, while a person's productive disposition relates to finding meaning in mathematical processes and believing in one's abilities. These aspects provide a framework for discussing children's knowledge, skills, abilities and beliefs in developing mathematical proficiency. To describe the different components of number sense, I also needed a more specific model to describe these skills and therefore apply the FoNS model for operationalising number sense (further described in Chapter 4).

2.1.3 Three related perspectives on number sense

The previously introduced perspectives on number sense can be categorised within the three related perspectives described as preverbal, foundational and applied number sense (Andrews & Sayers, 2015) or approximate, early and mature number sense (Whitacre et al., 2020). While preverbal or approximate number sense refers to a person's innate capability to compare quantities, foundational or early number sense is acquired through instruction and developed in children’s first year of schooling. Finally, mature or applied number sense relates to flexibility in number operations when applying mathematical knowledge and skills in a person’s everyday life. Because of my focus on children’s number sense development between the ages of 5 and 7, this thesis relates to an early number sense perspective on number sense. Consequently, most of the reviewed research relates to an early or foundational perspective on number sense, connected to the number sense knowledge children acquire and learn during preschool and the start of primary school.

Reviewing the number sense concept also reveals descriptions connected to broader frameworks for mathematical understanding and competence. Research related to the mature or applied perspective on number sense often involves older children or adults and includes descriptions related to flexibility and a deeper understanding of mathematics (Whitacre et al., 2020). Therefore, research on early number sense or FoNS has focused on the observable skills children learn in school. At the same time, FoNS is also linked to descriptions of number sense as interrelated knowledge involving flexibility and conceptual structures (Andrews & Sayers, 2015). Thus, I needed to connect the different perspectives and descriptions to enable the assessment of number sense.

2.1.4 The connections between the different perspectives

To describe children’s number sense at the start of primary school, I needed to separate the components represented through the FoNS model containing the number sense skills to be assessed. In this way, I look at the different constituents of number sense from an analytical point of view. At the same time, I also needed to consider a broader picture of real-world situations in which this competence is used, including the flexibility and connections between the components that comprise children’s number sense. In several of the papers in the thesis, I emphasise how technology can enable us to inform teachers about *how* children find the answer to assessment items rather than just showing *if* the children’s answers are right or wrong. Therefore, the applied number sense model is also connected to the mathematical proficiency framework, which illustrates the competence needed to perform mathematics. Together, these conceptualisations form the construct that I, from my perspective, call children’s number competence (illustrated in Figure 1). This conceptualisation of number competence includes more than the subdomains of early number competence (e.g. number, number relations, and number operations) as described in previous research (Devlin et al., 2022).

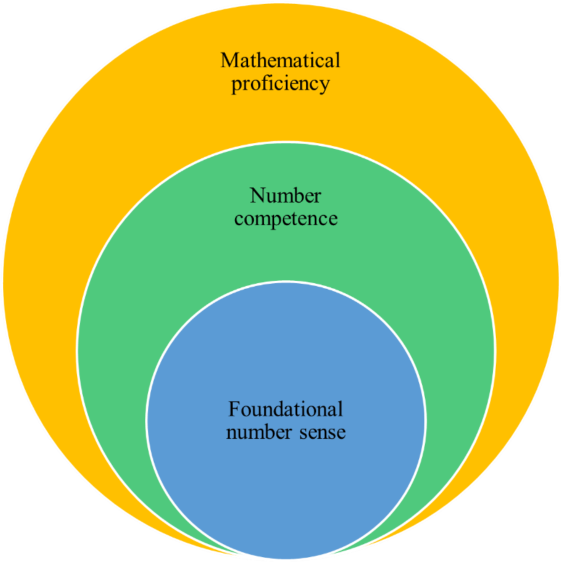


Figure 1. The interpretation of number competence in this thesis, comprising children's FoNS and a wider perspective of mathematical proficiency to illustrate the skills and competence needed to solve mathematical problems.

Further in this chapter, I will look more into research related to the first part of my research question concerning how a *specific digital resource can contribute to the formative assessment* of children's number sense.

2.2 Assessment practices for describing children's number sense

2.2.1 Formative assessment and equity

In this section, I will look into the current assessment practices for describing children's number sense using formative assessment and related tools. However, I first need to define central concepts when describing children's competence: formative assessment and equity.

Formative assessment, also known as assessment for learning, refers to the use of assessment to promote further learning, in contrast to summative assessment, which relates to evaluating a finished process or product (Black & Wiliam, 2018). There are different ways assessments can be used to promote learning. While formative assessment is meant to facilitate further learning processes, summative assessments can also be used formatively if the results of the assessment process are used to enhance further learning. Similarly, knowledge acquired from a formative assessment process may be used in summative ways by evaluating the assessment results as a finished product and not building on the results to facilitate further learning. Thus, there are nuances between the different assessment forms, depending on the intended use of the assessment (Bennett, 2011).

Formative assessment is strongly related to equity. While formative assessment involves assessment processes that provide opportunities for learning based on an individual's current competence, equity means that everyone has the same rights to be given these opportunities, regardless of who they are or where they come from. Equity in mathematics education refers to fairness concerning equal opportunities for all to have access to and achieve mathematical competence, to identify themselves as proficient in mathematics and to have the power to use mathematical knowledge in various contexts (Gutiérrez, 2012). In this sense, equity is not the same as sameness, supposing that everyone learns the same way and ends up in the same place. By informing learning to move the learning process forward, assessment for learning can contribute to equity in the mathematics classroom, where different children have different needs in their learning process (Heritage & Wylie, 2018). In this way, teachers can help children work at the "edge" of their knowledge to support further learning (Heritage & Heritage, 2013, p. 178). For the further development of assessment practices, new research on assessment in mathematics education should strive to improve equity and utilise the potential of technology (Nortvedt & Buchholtz, 2018). I have reviewed current assessment tools for describing young children's number sense to discover how this development can be facilitated.

2.2.2 Assessment format and context

An assessment can be delivered in various formats, from written assessments to individual assessment interviews. Different advantages and challenges come with different assessment formats, and different assessment tools or methods help us describe children's knowledge in various ways. The knowledge gained from the assessment process depends on the assessment's intended purpose and the format of the assessment. Written assessments can give an overview of the competence of larger groups of children but can be challenging to apply with young children who might not know how to read or write. Individual assessment interviews can provide valuable insights regarding individual children's strategies and possibilities for further learning (Clarke et al., 2011), but they can also be very time-demanding.

The assessment context might also influence how a child's competence is described. Observation of children's early number experiences through play and everyday activities is a way of assessing children's number sense in a more natural setting than a written test (Bergsmo et al., 2020; Wager & Parks, 2016). The context, both in terms of the situated information on the assessment problem and the environment in which the assessment takes place, can affect how children perform on an assessment. While the context of the problem itself can hinder and promote children's performance on a task (van Den Heuvel-Panhuizen, 2005), the situation in which the assessment takes place can also become an issue. Children might respond differently to tasks given at school than they do at home or tasks that involve topics they are interested in instead of tasks they cannot relate to.

2.2.3 Assessment tools in the Norwegian context

In the Norwegian context, the school system is primarily influenced by the idea of a "School for all", giving all children equal access to education in an inclusive environment (Imsen & Volckmar, 2014). Only a few paper-based tests are available to assess Norwegian primary-aged children's number sense (Lopez-Pedersen et al., 2021). One challenge in the Norwegian context is that the available assessment tools either provide specific information related to some of the previously described counting principles or measure a wide range of mathematical areas without describing children's competence and further learning potential.

"Alle teller" or "Everyone counts" in English was developed for the Norwegian education system to assess children's number sense through individual assessment interviews (McIntosh, 2007). One digital test has also been developed to playfully assess children's knowledge of numeracy, geometry and problem-solving through 18 tablet test items (ten Braak & Størksen, 2021). While both tests can provide valuable information about children's number sense development, they describe only a few of the previously described dimensions related to children's number competence. Additionally, the Norwegian national mapping tests in reading and numeracy are, as part of the Norwegian quality assessment reform, voluntary to perform for the individual school in year one and compulsory in year three of primary school (Norwegian directorate for education and training, 2023). These mapping tests were digitised in the spring of 2022. The national tests aim to identify children who need special attention and support. Therefore, they do not provide information regarding most children's number sense development.

2.2.4 Assessment tools in the Scandinavian context

In a broader Scandinavian context, the Swedish educational system has developed the assessment material "Hitta matematikken" or "Find the mathematics" in English. This material, which is mandatory at the end of the first year of school, was developed for teachers to evaluate children's early mathematical competence using research-based assessment activities adapted for group work or individual interviews performed by the teacher (National agency for education, 2019). In contrast, the lack of assessment instruments for evaluating young children's early mathematical knowledge has also been an issue in the Danish research context, which has caused the need to translate existing assessments from other countries (Sjoe et al., 2019). Similarly, there is a lack of assessment instruments that provide teachers with information about individual children's knowledge and further learning potential in the Norwegian context. This might lead teachers to translate the measures used in other countries into the classroom, creating challenges regarding the validity of the assessment. The lack of available assessment tools for evaluating young children's mathematical knowledge in some Scandinavian countries might be connected to the view on assessing young children's knowledge in the early childhood education context in Scandinavia. In this context, assessments are mostly performed if one is worried about an individual child's development (Urban et al., 2022).

2.2.5 Assessment tools in the international context

Internationally, several assessment tools have been developed to assess young children's early mathematical knowledge (Geary et al., 2009; Ginsburg & Baroody, 2003; Jordan et al., 2010; Malofeeva et al., 2004; van de Rijt et al., 1999). While these measures are used and affect the teaching and learning of mathematics in the primary grades, the provided information regarding the validity and reliability of attainment measures is often inaccessible, missing or of low quality (Breadmore & Carroll, 2021). In a recent systematic review of attainment measures in the United Kingdom, Breadmore and Carroll (2021) found 16 norm-referenced tests for primary-aged children that supported their criteria. Additionally, few reviewed tests report criterion validity that relates the test measures to real-life measures and does not significantly reflect how variance in test conditions can affect test performance. Generally, there is a lack of focus on the validity evidence of measures in mathematics education (Bostic et al., 2019; Krupa et al., 2019).

2.2.6 A need for new assessment tools and the potential in technology

The review of the available tests in the Norwegian, Scandinavian and wider international contexts points to several reasons why there is a need to develop new methods for assessing children's number sense. I relate this need to three arguments for developing such assessment methods: format, context and validity.

Concerning assessment format, digital assessments can ameliorate some of the disadvantages related to written assessments and individual interviews, such as accessibility and time. Technology has the potential to combine some of the advantages of both written assessments and individual interviews when providing opportunities to develop assessments that can be distributed to larger groups of children without being dependent on the children's reading and writing skills. Digital assessments have become acknowledged tools for gaining insights into children's mathematical knowledge and needs (Clements et al., 2021; Ginsburg, 2016).

Another argument for developing new methods for assessing children's number sense relates to the assessment context. Tests need to be adapted to a specific educational context. Using translated materials without adapting an assessment tool to the context in which it is to be used may threaten assessment validity. Additionally, many of the available assessment tools in Scandinavian and international contexts are assessments of learning rather than directed at providing formative feedback. The validations of these assessment tools are also very narrowly focused on psychometric properties, lacking a focus on further aspects of assessment validity (Breadmore & Carroll, 2021). These further aspects of assessment validity will be presented in Chapter 3.

Different teachers use the different tools described in this section in various ways to facilitate children's learning. How the information from various assessments is used depends on the individual teacher's assessment literacy, which I will review further in the next section.

2.3 The knowledge teachers bring to the assessment process

To address the part of my research question related to *how an assessment tool can contribute to the formative assessment of children's number sense*, I need to turn to the teacher's perspective and teachers' formative assessment practices.

Teachers use their knowledge to interpret children's assessment results to facilitate the further learning process. This knowledge is often referred to as pedagogical content knowledge (PCK). PCK can be described through the three dimensions of PCK as defined in the COACTIV framework by Krauss et al. (2008) and Baumert et al. (2010, p. 143): knowledge of tasks, children's ideas and representations and explanations. Together, these dimensions describe teachers' knowledge related to understanding the potential of mathematical tasks to facilitate children's learning and knowledge about students' existing beliefs and assessment of their prior knowledge. Additionally, PCK concerns teacher knowledge regarding supporting children's learning by using various representations and explanations related to the individual child's needs. While these aspects are all central to teachers' assessment work, they are also situated in a larger social context (Hodgen, 2011). Teacher knowledge is influenced by social structures in the classroom context, cooperation with colleagues and school culture. Therefore, it is crucial to consider the context of formative assessment when looking at teachers' interpretations of children's assessment results. In further descriptions of the knowledge teachers bring to the assessment process, PCK is a component of teachers' assessment literacy (Xu & Brown, 2016).

Assessment literacy has been defined in various ways, focusing on the competencies of assessment-literate teachers (Brookhart, 2011; Popham, 2009), which includes knowledge about multiple forms of assessments, how to use them and how to interpret results for formative assessment purposes. Further research has connected assessment literacy to various levels in the context of assessment (Xu & Brown, 2016). What constitutes assessment literacy can have different meanings across different educational and cultural contexts, resulting in the need for a model of assessment literacy in context (Pastore & Andrade, 2019).

The three-dimensional model of assessment literacy was developed to support teachers in making sense of their assessment literacy in relation to educational policy requirements and enhancing children learning (Pastore & Andrade, 2019). The interrelated dimensions of the model (Figure 2), situated in a socioconstructivist perspective, consist of a conceptual, praxiological and socioemotional dimension. The conceptual dimension includes the teacher's ideas of what assessment is and its purposes related to theories of learning, different forms of assessments and quality, in addition to analysing assessment data and communicating assessment results. How assessment works in teacher practice relates to the praxeological dimension of the model. This dimension comprises aspects of the teaching and learning process related to what is to be learned, how to learn and collecting evidence of student learning to adjust instruction adapted to the curriculum. The praxeological dimension also includes engaging teachers, parents and other interested parties with assessment information and communicating with children to regulate their learning process. Finally, the socioemotional dimension of the model incorporates the social practice of assessment related to communicating a shared conception of assessment practice in favour of children's further learning. In addition, the socioemotional dimension includes teachers' knowledge of their roles in the assessment process related to formal and ethical considerations, including the consequences of assessment and equity. Moreover, the socioemotional dimension includes negative emotions from a child's point of view that can influence the assessment process.

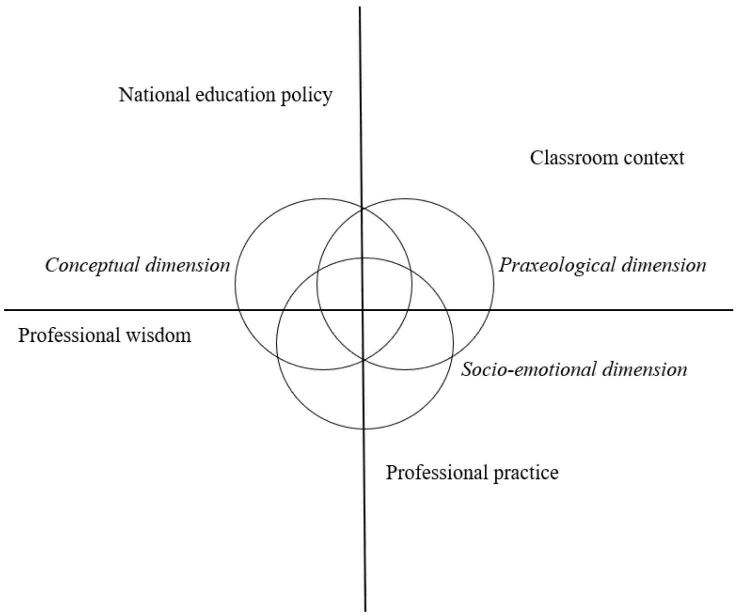


Figure 2. The three dimensions of assessment literacy presented by Pastore and Andrade (2019, p. 135).

2.4 Chapter summary

In this chapter, I have reviewed the relevant literature related to the main concepts and practices in the thesis: number sense, formative assessment tools and teachers' assessment literacy. I have focused on the number sense concept and looked at how it has been viewed through different definitions and components. Furthermore, I have looked at different assessment tools and how they can be used for the formative assessment of children's number sense.

The conceptual frameworks that frame the thesis were identified in the literature review. To enable the investigation of how a specific assessment tool can contribute to the formative assessment of children's number sense at the beginning of first grade, I have applied the FoNS model to define eight interrelated number sense categories (Andrews & Sayers, 2015). Furthermore, I have added subitising as an additional category. The content of each FoNS category and reasons for adding subitising as one of these are described in more detail in Chapter 5. 1. 1. Further operationalisation of the construct is described in Chapter 4. 4.

Finally, I have introduced perspectives on teachers' practice, focusing on assessment literacy to describe how teachers can use information from formative assessments to facilitate children's further number sense learning. The presented framework on teachers' assessment literacy (Pastore & Andrade, 2019) makes it possible to look further into the consequences of the assessment process. Teachers' PCK, as part of their assessment literacy, is situated in a wider social context in which their surroundings influence teachers' interpretation and use of assessments. The described dimensions of teachers' assessment literacy also relate to assessments' instructional validity, which is presented in the next chapter. In the next section, I will also describe how the assessment of number sense implies two different ways of looking at the learning process and how validity theory may inform assessment development.

3 Theoretical frameworks

3.1 The process and product of learning

A central goal of this thesis is to improve children's opportunities for learning number sense at the start of primary school. Simultaneously, using information from an assessment process to facilitate further learning (i.e. formative assessment) does not entail any specific view of what happens in the learning process (Wiliam, 2018). Different views on the learning process entail different perspectives on children's learning opportunities. I will further describe how using a specific digital assessment tool to improve children's number sense can relate to two different perspectives on learning before I relate these perspectives as the process and product of learning.

Operationalising number sense through nine different categories, presented as tasks in a digital assessment tool that is further used to evaluate children's number sense, implies describing an individual's separately constructed reality. In this way, children's number sense learning opportunities depend on their mental schema (Piaget, 1970) of number sense and each child's individual development. In a constructivist frame, number sense would only exist as mental models that vary from person to person, and the learning opportunities would thus depend on these individual mental schemas.

At the same time, number sense can be seen as a conventional object abstracted from mental models in social practice. The digital assessment tool is developed for formative assessment of children's number sense to inform teachers about their children's knowledge. Providing teachers with information meant to promote further learning would indicate a sociocultural view of learning, advancing the child's developmental level within the proximal zone of development (Vygotskij, 1978). Focusing on both the development and use of the tool makes it challenging to define this thesis within one specific theoretical framework of learning, as development and use imply different views on the learning process.

Investigating children's number sense learning opportunities using a digital assessment tool involves a contradiction and tension between the sociocultural and constructivist perspectives on learning. This tension can be resolved by taking different perspectives at different project stages. Attending to number sense as a product by investigating what children have learned about numbers when they start school would imply a constructivist perspective. Shifting attention to the children's further learning process, studying how teachers might use the information about their children's number sense would involve sociocultural perspectives. At the same time, the focus of this project will lie within a product orientation of number sense, describing what children have learned at one specific point and how this product can inform further learning processes. The process and product of children's number sense are described and evaluated differently. Simultaneously, the process of learning and the product of the assessment of number sense are clearly related but need to be explained in distinct terms.

3.2 Validity theory

Validity theories in educational measurement were initially developed to evaluate the intended use and interpretations of assessment scores, presenting principles, practices and types of evidence needed to evaluate these interpretations (Moss et al., 2006). Validity theory can be described as a scientific enquiry (Cronbach & Meehl, 1955; Messick, 1989), as a practical argument (Kane, 2016; Kane et al., 1999) or with a broader focus on the consequences of assessment practices (Moss, 2013; Moss et al., 2006). In the context of assessment development and measurement of children's number sense, I look at validity theory as a scientific enquiry, presenting the validity evidence of ENSA. In this way, validity refers to how theory supports the inferences made from test scores and how the evidence strengthens interpretations of test scores (Wolfe & Smith, 2007b). At the same time, I also look at the consequences of the assessment process from a teacher's point of view. For this, I need a different framework to inform the consequential or instructional aspect of validity (Pellegrino et al., 2016).

The application of Rasch measurement theory entails looking at validity theory through the descriptions of a process of instrument development oriented towards identifying validity evidence. To describe this process and related validity evidence, Wolfe and Smith (2007) adapted Messick's (1995) six aspects of validity with criteria presented by the Medical Outcomes Trust (1995), presenting in total eight validity aspects: 1) content, 2) substantive, 3) structural, 4) generalisability, 5) external, 6) responsiveness, 7) interpretability and 8) consequential. In Chapter 5.6, I describe how the validity aspects are addressed in the papers in the thesis.

Content validity refers to the relevance and representativeness of the content of the items, which, in the context of this thesis, are based on the description of the categories in the FoNS framework and descriptions from previous assessment instruments. The technical quality of the items is also an essential part of content validity. Evidence of content validity can be presented by describing instrument purpose, test specifications, item development, expert reviews and analysis of item technical quality.

The substantive aspect of validity refers to the relationship between the theoretical construct of number sense, children's cognitive response processes and the observed consistencies among item responses. To evaluate the evidence of an assessment's substantive validity in a Rasch measurement context, one can look at the response structure of the participants by analysing the person fit and item difficulty hierarchy.

In this PhD thesis, the structural aspect of validity refers to how the scoring structure relates to the structure of the FoNS model, which represents the construct domain. Evidence of the structural aspect of validity can be evaluated through a dimensionality analysis, comparing the structure of a participant's responses to the structure of the construct domain.

The generalisability and external aspects of validity refer to external meanings. Generalisability relates to the meaning of the measures and interpretations across contexts and can be addressed through evidence of assessment reliability. External validity refers to the measure's relation to external measures of the same construct in addition to the relevance and usefulness of the measures. Evidence for external validity can be presented through group comparisons or by looking at changes in individual performance.

The responsiveness aspect of validity relates to the sensitivity or capacity of the assessment tool to detect changes after an intervention. Evidence of responsiveness can be presented by analysing the tool's capacity to describe individual development.

Finally, the interpretability and consequential aspects of validity entail looking at validity from a broader perspective. Interpretability refers to the degree to which the meaning of measures is communicated to those who need to interpret them. From this perspective, qualitative meaning must be assigned to quantitative measures to be understandable to individuals unfamiliar with the measurement context. Evidence of the interpretability aspect of validity in our context relates to the degree to which teachers can make sense of children's assessment results. Consequential validity implies the consequences of measures and the implications of courses of action related to bias and fairness. The consequential aspect of validity can be addressed by evaluating the scoring procedure and the consequences of the assessment process. I needed to apply different frameworks to evaluate the interpretability and consequential aspects of ENSA from a broader perspective, not only related to the Rasch measurement context.

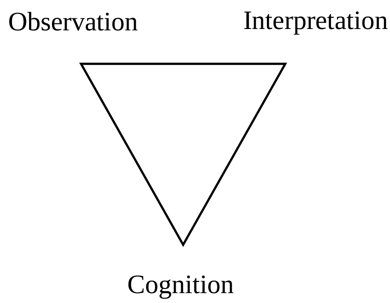


Figure 3. The assessment triangle (Pellegrino et al., 2001, 2016)

The assessment triangle (illustrated in Figure 3) provides a framework for analysing and designing assessments by representing three key elements that underlie assessment (Pellegrino et al., 2001): cognition, observation and interpretation. The cognition element relates to children's cognition and learning in the assessment through theories and models that represent the subject matter domain. The observation element indicates the assumptions and principles about the evidence of children's competence according to children's learning. The critical element interpretation indicates making sense of the assessment's evidence according to the assessment's purpose and relates to inferential validity and instructional implications. The three elements of the assessment triangle are crucial for the arguments regarding the assessments' cognitive, instructional and inferential validity. For an assessment to be valid, all three vertices of the triangle must synchronise (Pellegrino et al., 2016). Looking specifically at teachers' use of assessment results, the cognitive validity of the assessment refers to teachers' expectations of children's performance. In this context, looking at the applied number sense (FoNS) framework related to teachers' expectations will be relevant. The cognitive aspect of validity can be compared to the substantive aspect of validity in the previously described validity approach. Investigating the instructional validity, parallel to the consequential validity, of the assessment implies looking at how teachers understand and use the results provided by the assessment. It is central to consider teachers' use of assessment results and how differential decisions and actions are supported (Pellegrino et al., 2016). From teachers' perspectives, the inferential validity aspect relates to teachers' previous experiences with assessments describing children's knowledge and might thus, to some extent, relate to the previously described interpretability aspect of validity. The way the framework by Pellegrino et al. (2001, 2016) and the approach by Wolfe and Smith (2007) relate is described in Figure 4.

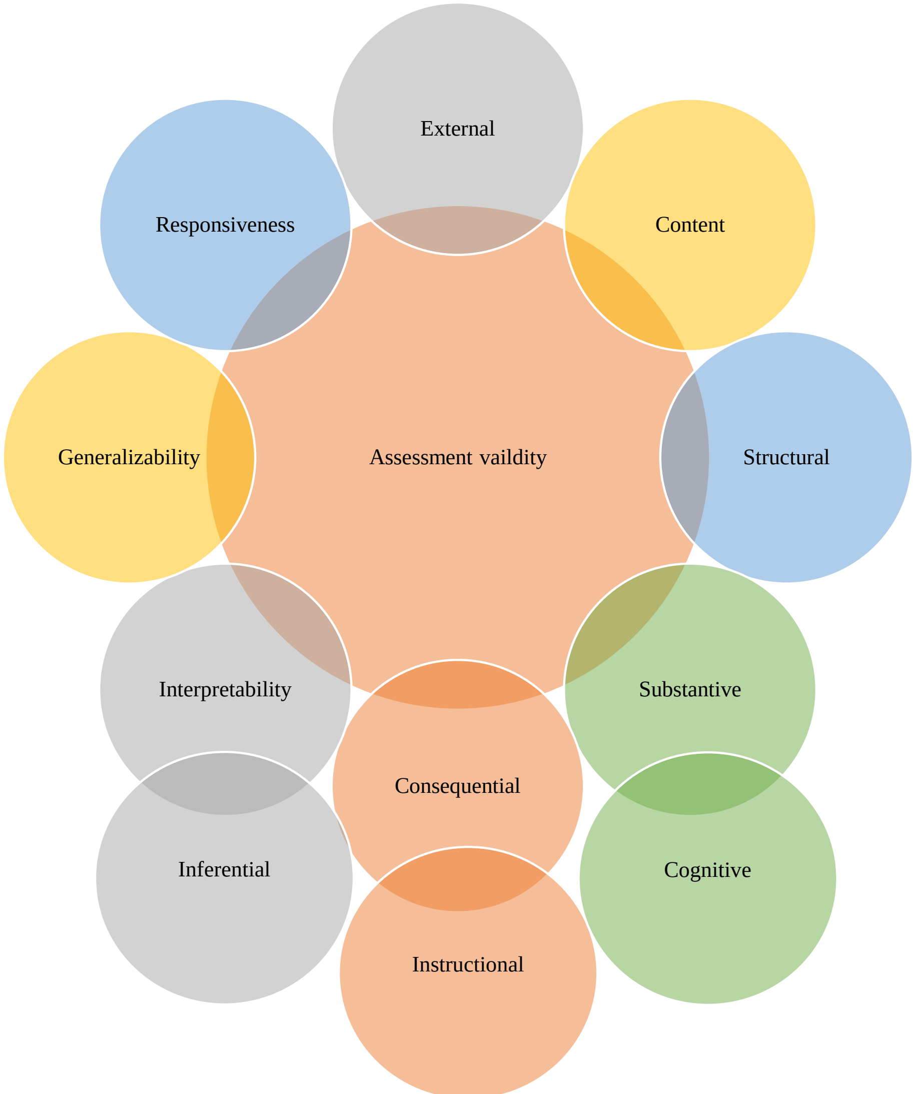


Figure 4. The eight validity aspects as described by Wolfe and Smith (2007) – content, substantive, structural, generalisability, external, responsiveness, interpretability and consequential – and how they relate to the three aspects presented by Pellegrino et al. (2001, 2016) – cognitive, instructional and inferential.

3.3 Chapter summary

In this chapter, I have described how two views on the process and product of learning need to be applied at different stages of investigating the development and formative use of an assessment tool. I have also illustrated how validity theory in a Rasch measurement context can provide insights into various aspects of assessment validity. However, I also need other frameworks to describe the instructional aspect of validity. The next chapter describes how the different views on learning and validity theory connect to Rasch measurement theory in a design-based research context.

4 Methodology

Investigating how a specific digital resource (ENSA) can contribute to the formative assessment of children's number sense at the beginning of first grade requires a combination of different ontological and methodological perspectives. Figure 5 illustrates how a pragmatic paradigm facilitates the combination of qualitative and quantitative perspectives within a design-based research methodology. In this way, I can investigate the development and use of ENSA to identify how the tool can contribute to formative assessment practices. In the following chapter, I will detail the applied methodology that enables this investigation. My main focus will lie on the quantitative perspectives, although I will also connect the overarching perspectives that enable qualitative thematic analysis. Further details regarding the methods applied in the thesis are presented in Chapter 5.

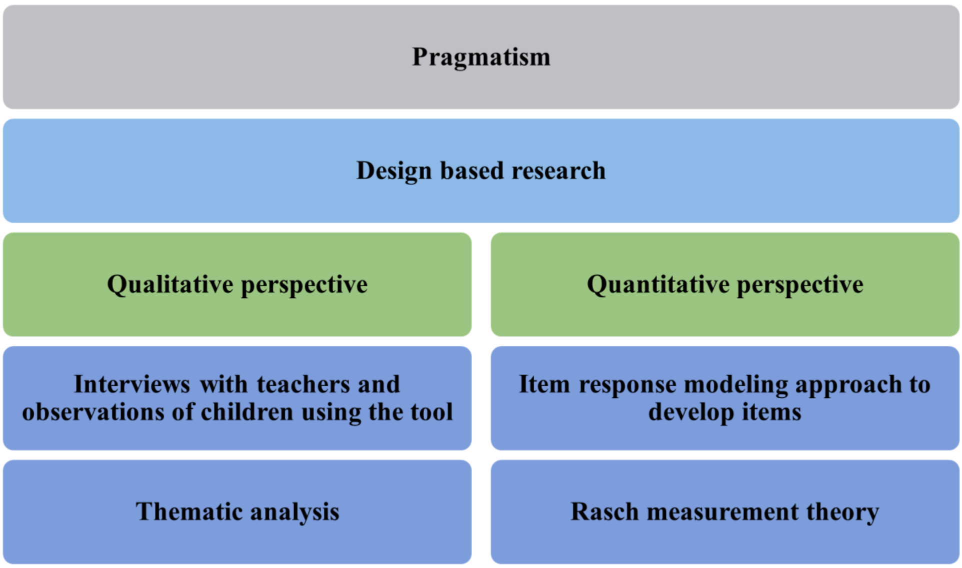


Figure 5. The overarching pragmatic paradigm enables the design-based research process for developing ENSA, combining qualitative and quantitative research methods and analytical perspectives to investigate the development and use of the tool.

4.1 A pragmatic paradigm for combining different research methods

Investigating both the development and use of ENSA requires different research methods and, consequently, different ontological perspectives. To describe how this specific digital assessment tool can be used for formative assessment purposes, I combine qualitative and quantitative research methods within a pragmatic paradigm.

Qualitative and quantitative methods are based on conflicting paradigms and can be described as two ends of the research continuum (Yilmaz, 2013). At one end of this continuum, quantitative research is based on an objective epistemology, searching for objective truth and describing a reality independent of the object of study. Qualitative research, at the other end of the continuum, is based on a constructivist epistemology that studies constructed, dynamic reality through detailed descriptions of the process of meaning-making and understanding. Investigating the development and use of ENSA involves objective questions related to quantitative terminologies, such as reliability and validity. Looking more closely at how teachers use the assessment tool involves a qualitative inquiry into how individuals understand the assessment results. Combining qualitative and quantitative methods makes it necessary to highlight which arguments are put forward, justify these choices and clarify epistemological issues regarding what it means to attain knowledge from these perspectives (Biesta, 2010).

Pragmatism is often introduced as a natural paradigm when combining qualitative and quantitative methods (Johnson et al., 2007; Morgan, 2007; Weaver, 2018). Pragmatism is not based on a specific ontological or epistemological stance and can therefore combine different research approaches. Within a pragmatic paradigm, I can apply different research methods to investigate the development and use of ENSA.

4.2 The design-based research process

In this thesis, I use a design-based research approach to understand how ENSA can promote the formative assessment of young children's number sense. Brown (1992) and Collins (1992) first introduced design experiments as an interventionalist methodology in which the goal was to engineer and simultaneously investigate educational innovations and further develop design theories for future innovations. Accordingly, the design-based research process implies designing and testing innovations and identifying elements that work (Bakker & van Eerde, 2015). Designing ENSA and evaluating the validity evidence of the tool is, in this way, connected to design-based research through the development of an assessment tool and the identification of how this tool can contribute to the development of further assessments.

While pragmatism can facilitate the connection to an interventionalist design (Biesta, 2010), design-based research has been described as a bootstrapping, interventionist methodology (Cobb et al., 2015). A problem in the design-based research literature is the lack of discussion of epistemological issues when combining qualitative and quantitative approaches (Walker, 2011). However, design-based research data are mainly qualitative (Cobb et al., 2015; Godino et al., 2013). Still, a combination of qualitative and quantitative approaches can, as in my case, be necessary (Middletion et al., 2008).

This thesis comprises the start of a design-based process investigating the development and use of ENSA to describe first-grade children's number sense. Through iterative cycles, design-based research involves studying the development of particular forms of practice, programmes or artefacts (Cobb et al., 2015). The process involves the operationalisation, design, testing, investigation and evaluation of the assessment tool (Figure 6).

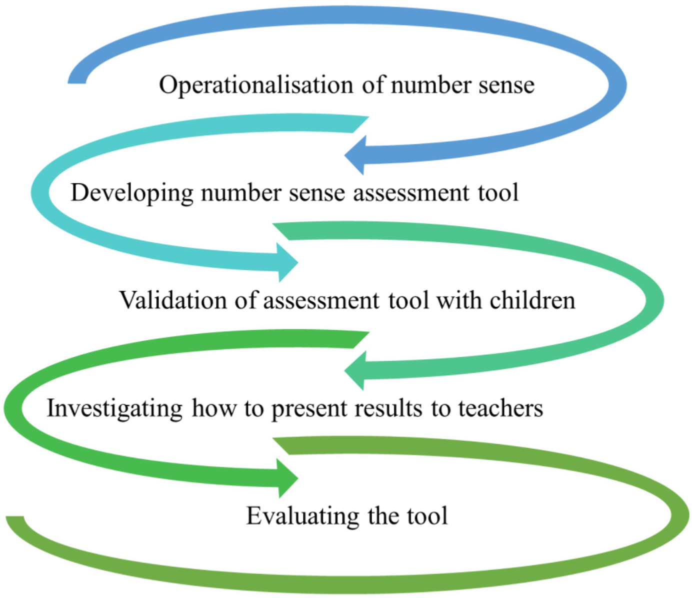


Figure 6. The design-based research process of ENSA: operationalising number sense and further developing, validating, investigating and evaluating the assessment tool

As described in Figure 6, the design-based research process started with the operationalisation of number sense. Furthermore, the design-based research process involved developing the assessment tool. This development process, resulting in the digital assessment tool ENSA, will be described in more detail in Chapter 5.1. Additionally, Chapter 5 will provide details into the further process – the data collection with 5–6-year-old children who carried out the assessment. Next, the design-based research process involved examining how ENSA could function for formative purposes by presenting children's assessment results to teachers. Finally, this thesis comprises the evaluation of the assessment tool before continuing the design-based process to refine the tool for further use by teachers in primary schools.

4.3 Reasons for applying Rasch measurement

When considering measurement, what comes to mind first is often the quantification of spatial properties, such as length, area and volume. However, in the social sciences, one can also use some of these properties to enable the measurement of an abstract construct, such as number sense. Measurement implies abstracting qualitative observations from the real world (Thurstone, 1959) and exchanging the observed data for inferred meaning. In this way, one assigns numbers to categories where the numbers have certain properties, checks that the assignment was successful and makes use of the measurements for the purpose of summarising responses (Wilson, 2005). In this thesis, I apply the FoNS model to define categories of number sense. Based on the descriptions of each category, I designed tasks that served as indicators of children's number sense through their answers on ENSA. Further, I applied Rasch measurements to make sense of the data.

Applying the Rasch model enabled me to overcome the issues I encountered when measuring young children's number sense. More specifically, Rasch measurement can be applied to overcome issues related to producing linear measures, overcome problems related to missing data and detect misfit and parameter separation (Wright & Mok, 2004).

First, the dichotomous Rasch model (Rasch, 1960) enabled me to produce linear measures of the children's performance that are typically used in knowledge assessments with two levels of performance and item types, such as multiple choice. Furthermore, a problem one might encounter when assessing young children's knowledge is the problem of missing data. Rasch analysis allowed for the comparison of children's performance, although not all children had completed the same items in the assessment. Additionally, in the development process of ENSA and the further analysis of the data, I had to detect items that did not function to the same extent as the other items to measure children's number sense or individual children's responses that, for some reason, did not respond to the items in the same way as the majority of the children. In this way, Rasch measurement enabled me to make measures that were not dependent on specific items or individuals where I could detect and evaluate misfitting items or individual responses.

4.4 Item response modelling approach to measure number sense

To measure an abstract concept, such as number sense, I needed to operationalise the construct. An item response modelling approach (Wilson, 2005) to instrument development enabled the construction needed to develop measures of children's number sense, identify central components and analyse how the different measures relate. The item response modelling approach involves four building blocks: 1) a construct map, 2) item design (explained further in Chapter 5.1), 3) outcome space and 4) a measurement model.

The construct map includes a definition and description of the construct that is to be measured – in the case of this thesis: number sense, with an ordering of items that signalises whether there is more or less of the construct. The map representing a construct of the theoretical object number sense is also a simplistic illustration that does not display the flexibility and relations between the indicators. Figure 7 displays an example of a construct map developed in the early stages of the ENSA design process. When ordering the list of indicators of number sense on the construct map, I chose 18 of the developed items at that time. Two different items from each FoNS category were chosen to order responses to items from low to high number competence. The items that were expected to be the most difficult were placed at the top of the map, and the easiest items were placed at the bottom. This was done before any data collection was conducted. The results presented in Paper IV confirmed that item NP7 was one of the most difficult items in ENSA, thus verifying the initial hypothesis about this item.

The outcome space involves categories connected to the scoring of items (Wilson, 2005). The children's answers to ENSA were scored as right or wrong, giving one or zero points to each item. The Rasch measurement model enabled the analysis of the data, which is further presented in Chapter 5.5.

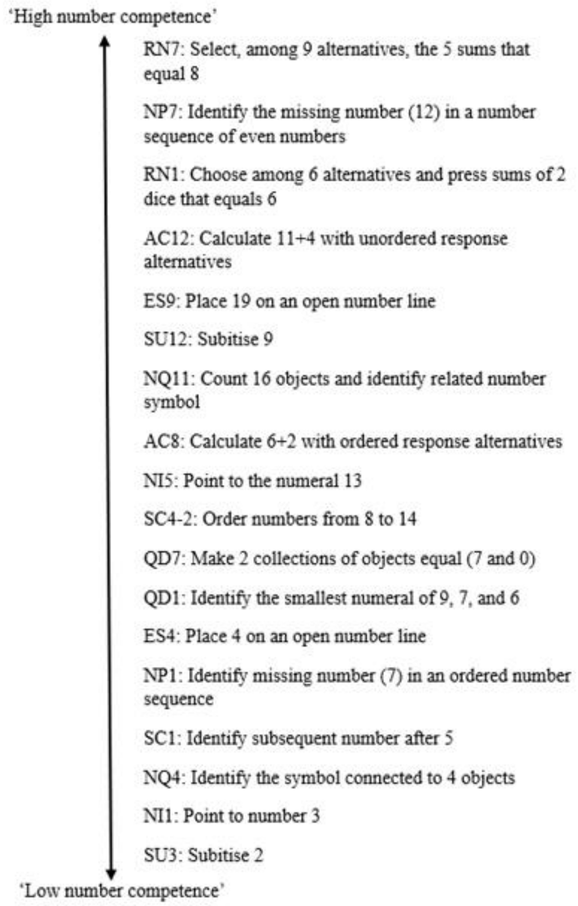


Figure 7. A construct map from the development process of ENSA, ordering items from low to high number competence and increasing with the correct answers on difficult items.

4.5 Rasch measurement theory

The children's answers to ENSA were scored as right or wrong and analysed with Rasch measurement theory, more specifically, the dichotomous Rasch model (Rasch, 1960).

4.5.1 Properties of Rasch models

In the same way that a ruler has the required properties for measuring distance, Rasch models have the properties of sufficiency, separability, specific objectivity and latent additivity required for measuring latent variables such as number sense (Rost, 2001). The property of sufficiency implies that the total responses for persons and items are sufficient statistics for item and person estimation and contain all information about a person's level. In the context of this thesis, sufficiency implies that the collected responses to ENSA comprised sufficient statistics for conducting the necessary estimations to describe children's number sense.

Separability is a property of Rasch models, which states that it is possible to estimate item parameters without estimating or knowing the distribution of the person parameters and estimating person parameters without knowing the items. The separability property was a clear advantage when assessing young children's number sense – not all children answered all the items in ENSA. In this way, although the children did not answer all the items for various reasons, the model could estimate the person parameters.

The third property, specific objectivity, is required for invariant comparison, stating that an item's difficulty or a child's estimated number sense must be independent of the items used for measurement or the individual respondents. This is also related to test validity, which is further described in Chapter 3.2. The test results used in raw score forms lack the property of specific objectivity.

The final property of Rasch models, latent additivity, is also related to specific objectivity. If the property of specific objectivity holds, a representation of the model can connect the person and item parameters by addition or subtraction (Rost, 2001).

4.5.2 The dichotomous Rasch model

The Rasch model is a probabilistic measurement model that provides interval-scale measures of both item difficulty and person skill on the same measurement scale in units of logits (Wright, 1977). In the dichotomous Rasch model, $\boldsymbol{X_{vi}}$ is a dichotomous stochastic variable, implying that $\boldsymbol{X_{vi}}$ only has two possible outcomes: 0 or 1. The children who carried out ENSA received one point for a correct answer and zero points for a wrong answer; thus, all items were scored dichotomously. The probability that $\boldsymbol{X_{vi}}$ gets the outcome 1 and that person v scores 1 point on item i is given by the following function:

$$\boldsymbol{P}\,\{\boldsymbol{X_{vi}} = \boldsymbol{1}|\boldsymbol{\beta_v}, \boldsymbol{\delta_i}\} = \frac{\boldsymbol{e^{(\beta v\,-\,\delta i)}}}{\boldsymbol{1} + \boldsymbol{e^{(\beta v\,-\,\delta i)}}}$$

where β_v denotes the person's skill and δ_i denotes the item difficulty. The probability that person v scores one point on item i then depends on the difference between the person's skill β_v and the item difficulty δ_i. The probability that a person gets one point on an item with the same difficulty as the person's skill level is $\boldsymbol{P}\,\{\boldsymbol{X_{vi}} = \boldsymbol{1}|\boldsymbol{\beta_v} = \boldsymbol{\delta_i}\} = \boldsymbol{0,5}$. The parameters β_v og δ_i are calculated numerically in an iterative process in the program Winsteps (Linacre, 2017).

4.5.3 Rasch model requirements

Applying any Rasch model implies some requirements for the data (Smith & Wind, 2018). The two main requirements are local independence and unidimensionality. Local independence indicates that a participant's responses to one item should be statistically independent of responses to any other items (Hambleton & Swaminathan, 1985). This means that a child's response to one of the items in ENSA should not be dependent on previous or further answers to items. Unidimensionality means that all items in the instrument measure a shared dominant construct. In the case of measuring children's number sense with a basis in the FoNS model, this means that the different categories of the model together form a single measure of children's number sense. The case of dimensionality and the complexity of the number sense construct have been described in Chapter 2.

4.6 A thematic analysis of the qualitative data

As illustrated in Figure 5, the qualitative data from the design-based research process of ENSA were analysed by applying thematic analysis. Since thematic analysis is not bound to a specific epistemological stance (Braun & Clarke, 2006), it was possible to apply this method of analysis within a pragmatic paradigm. The thematic analysis was conducted to identify repeated patterns of meaning in the teachers' reflections on the children's assessment results, with the teachers' statements as the unit of analysis. In Chapter 5.5.2, the different stages of the thematic analysis are presented.

While it is possible to go into the thematic analysis process with different relations to theory, it is important to make the initial theoretical assumptions clear (Braun & Clarke, 2006). Theories connected to instructional validity (Chapter 3.2) were central to the initial stages of qualitative data collection in this thesis. However, I needed further perspectives from the literature on assessment literacy (Chapter 2.3) to make sense of the data.

4.7 Chapter summary

In this chapter, I have described how the design-based research process of ENSA connects to the application of Rasch measurement theory to model and operationalise the number sense construct. I have also described how pragmatism serves to combine qualitative and quantitative approaches in the design and evaluation of ENSA. In the next section, I will present the development process of ENSA and how the tool was used to collect the data.

5 Methods

I have applied qualitative and quantitative methods to investigate how ENSA can contribute to the formative assessment of first-grade children's number sense. While the development of the tool is a major part of the thesis, it is not described in detail in any of the included papers. Therefore, a substantial part of this methods chapter is dedicated to describing the development of ENSA. Still, many details regarding the choices made during the process cannot be described within the frames of this thesis. Consequently, I present only the most central aspects of the development process and reflect on the methods applied in the research, including data collection, participants, validity and ethical considerations.

5.1 Assessment development

ENSA was developed over one year, starting in the fall of 2019. A timeline of the development is described in Figure 8. After an initial literature review, items for each FoNS category were developed during three phases with expert reviews and pilot studies. In the following chapter, I will present how the literature review, expert reviews and pilot studies informed the item design and considerations made in the final assessment tool.

Figure 8. The phases of the development of ENSA, starting with Phase 1 in the fall of 2019 and resulting in the main data collection in the fall of 2020.

5.1.1 Literature review

I started the item development by reviewing existing research on tasks for measuring children's understanding of number sense categories in the FoNS model. Furthermore, I considered the content of each category and related items in alignment with the The knowledge promotion reform (Norwegian directorate for education and training, 2020). To identify items that would serve the intended purpose, I applied specific criteria for selecting items, ensuring that they 1) had a record as assessment items, 2) could be scored and developed into measurement items, 3) could be digitised and preferentially combined with some interactivity and 4) represented the abilities related to one of the FoNS categories. There was some flexibility in using these criteria for categories where few items were found in previous research. Item suggestions for all categories were gathered and introduced in the first expert review in Phase 1 (Figure 8).

5.1.1.1 Number identification

To assess the first FoNS category, number identification, determining whether a child could recognise a number, know its vocabulary and meaning and identify a symbol from a set of symbols, I looked at tasks from previous research, specifically the number set test and the point to x task (Geary et al., 2009; Wynn, 1992). The items I was inspired by had been used in different ways with only number symbols, number symbols combined with sets and only sets of objects. According to the FoNS model, there was a need to isolate the child's comprehension of the number symbol and its vocabulary. Therefore, I focused on items presenting children with number symbols in this category.

5.1.1.2 Systematic counting

For the systematic counting category, deciding if a child could count flexibly to 20 and back and upwards and downwards from a given point and determine how numbers have their given place on a sequence of numbers, I developed items inspired by Malofeeva et al. (2004). I also investigated possibilities in describing the flexibility aspect of a child's counting related to the counting judgement task (LeFevre et al., 2006). The systematic counting category also involved ordinality (Andrews & Sayers, 2015). These items were inspired by items where children were to point to, remove or insert objects in specific positions in a line (Malofeeva et al., 2004).

5.1.1.3 Number and quantity

I developed items inspired by three tasks to assess children's knowledge of numbers and quantities, including cardinality and the correspondence between symbols and quantity. In the items inspired by the point to the number task (Malofeeva et al., 2004), the children were asked to find a number symbol corresponding to a given quantity and a quantity corresponding to a given number. In the items inspired by the frequently used give-n task (Le Corre & Carey, 2007; Malofeeva et al., 2004; Wynn, 1992), the children were to drag and drop a given number of objects onto a specific area. Finally, I included items involving counting objects (LeFevre et al., 2006) in the number and quantity category. For these items, the children were asked to arrange randomly placed objects on the screen to be able to count how many there were.

5.1.1.4 Quantity discrimination

In the category of quantity discrimination, I needed to investigate the children's understanding of quantities and their capability to compare quantities using vocabulary such as larger, smaller, more than and less than. Therefore, I developed items inspired by tasks involving comparing sets (Le Corre & Carey, 2007; Malofeeva et al., 2004) and comparing numbers (Malofeeva et al., 2004).

5.1.1.5 Arithmetic competence

In the FoNS model, simple arithmetic competence involves the transformation of small quantities using addition and subtraction. When designing the items for this category, I took inspiration from word problems in other assessments (Ginsburg & Baroody, 2003; Ginsburg & Pappas, 2016). I also used a framework for developing different problem types for addition and subtraction: change and combine (sum) as well as compare and equalise (difference) (Carpenter & Moser, 1984). Different problem types were included in ENSA based on information from previous studies, indicating that children are highly capable of solving addition and subtraction tasks by modelling the situations in the problem (Carpenter et al., 1993). The sum tasks were composed in two ways. Either with one initial quantity and an action that causes a change (join) or with two initial quantities that may be considered separately or as a part of a whole. Difference tasks involve comparing two quantities to determine the difference between them. Additionally, equalise problems include an additional action that is to be performed to make the two sets equal.

5.1.1.6 Subitising

Subitising was added as a category in the assessment during Phase 2. Subitising is often referred to as the ability to quickly determine the size of a set of objects without counting the objects explicitly. Two different types of subitising are perceptual subitising and conceptual subitising. Perceptual subitising implies determining the quantity without any other mathematical processes. In contrast, in conceptual subitising, one would structure the objects mentally and use mental arithmetic to determine the size of the set (Clements et al., 2019). There were two main reasons for adding subitising as a separate category in ENSA: 1) subitising is highlighted as an essential aspect of developing number sense (Clements et al., 2019; Sayers et al., 2016) and 2) subitising is central to a child's further arithmetical development (Wilkins et al., 2022).

The subitising items were inspired by typical flashcard tasks (Clements et al., 2019). Additionally, the arrangement of the objects was considered because this is essential for a child's ability to subitise. Given developmental progression, objects in lines or rows are more accessible to subitise than more scattered presentations (Clements & Sarama, 2021). Based on these descriptions, four different types of subitising patterns were included in ENSA (Figure 9).

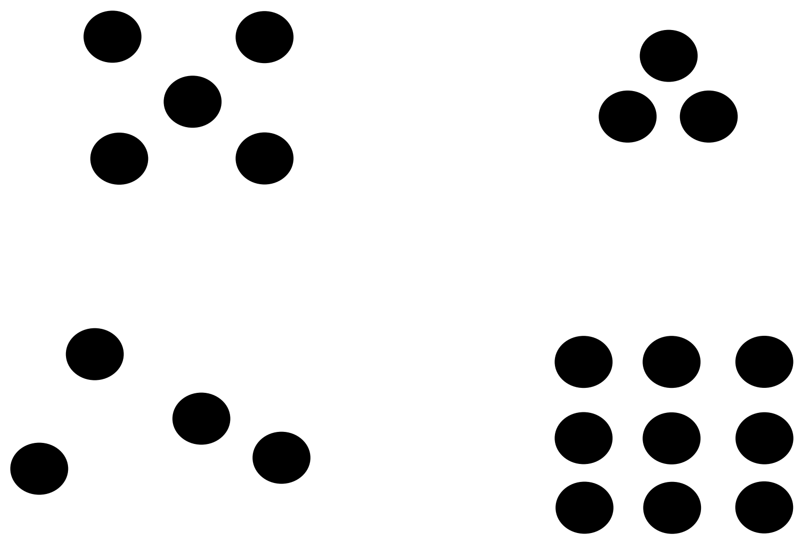


Figure 9. The four types of subitising patterns included in ENSA: Dice, pyramid, scattered and vertical lines.

5.1.1.7 Representing number

In FoNS, representing numbers involves children's understanding of different ways to represent numbers. The different representations of numbers are described as number lines, manipulatives and fingers (Andrews & Sayers, 2015). Partition (part-whole) is also regarded as a number representation in the FoNS model. The development of the first representing number item was inspired by a task described by Andrews and Sayers (2015, p. 264). The task was performed in a classroom context with a teacher where the children were given six pebbles to arrange in two different groups. A similar version of this task had a record as an assessment item (Clements et al., 2008) and the task could also be adapted to the digital format. Further adjustments of this item are presented in Chapter 5.1.5.3.

5.1.1.8 Estimation and number patterns

The final two FoNS categories to review the literature were estimation and number patterns. Estimation is described in the FoNS model as estimating the size of an object or a set and understanding different representations, such as an empty number line. For this category, I looked into previous research on number line estimation (Geary et al., 2009; Siegler & Booth, 2004). Awareness of number patterns includes identifying a missing number. Items in the number patterns category were developed from tasks with missing number patterns (Lembke & Foegen, 2009). I also investigated the possibilities in patterning tasks (Collins & Laski, 2015). All the items collected from the reviewed literature that fit the criteria were presented in the expert review in Phase 1.

5.1.2 Expert review

As illustrated in the timeline in Figure 8, expert reviews of the assessment items were carried out in Phases 1 and 2 after the initial literature review and the second pilot study. The expert reviews involved analysis of the item content related to the specific number sense categories, item design, item difficulty, context and instructions. The participants in the expert reviews were experts in the field of assessment in mathematics education, language development and children's mathematical understanding. Discussions and informal meetings were also held throughout the process.

5.1.3 Master students

In addition to the expert reviewers, there were six master students involved in the project. Gaarden (2020) and Søberg (2020) designed their own number sense assessment items and investigated children's strategies, and Ellefsen (2021) compared children's performance on paper versus digital test forms. Schjølberg (2021) looked at children's strategies when solving assessment items in two of the categories in ENSA, and Holvik (2022)investigated the teacher's perspective on children's assessment results. Finally, Neuwirt (2021) developed and researched the role of the graphical design, including setting the assessment in a farm context, the user interface and the choice of colours, shapes and sizes of objects adapted for the participants' ages. She also designed the blue arrow to send the children to the next item in ENSA (see Chapter 5.1.5.1) (Neuwirt, 2021). All six students contributed to the development of ENSA, from item design to teachers' use of the tool.

5.1.4 Pilot studies

Pilot studies were carried out in all three phases of the development process (see Figure 8), and items for the eight FoNS categories were developed during the three phases. In Phase 1, items for the first five FoNS categories were investigated: 1) number identification (11 items), 2) systematic counting (7 items), 3) number and quantity (10 items), 4) quantity discrimination (7 items) and 5) representing numbers (2 items). After ensuring that these items functioned as expected, arithmetic competence and subitising items were developed and tested in Phase 2. In Phase 3, items for all FoNS categories, in addition to subitising, were evaluated.

The pilot study in Phase 1 was conducted with 32 children in the first year of primary school in the spring of 2020. Six of these children carried out the assessment as part of individual interviews to enable the identification of qualitative aspects related to the functionality of the assessment tool. Additionally, individual interviews were conducted to investigate the possibility of including other assessment items in systematic counting and number patterns. For the items in the systematic counting category, the expert review in Phase 1 questioned to what extent the items developed at that time could inform teachers about the children's understanding of counting and their flexibility in applying different counting strategies. Therefore, the children carried out the counting judgement task (LeFevre et al., 2006), in addition to a patterning task (Collins & Laski, 2015), to see if these tasks were suited for the age group and possibly could be included in the assessment. During the interviews, the qualitative impression of the counting judgement task was that it could provide information about children's flexibility in the counting process related to counting in steps and from arbitrary positions in the counting sequence. Unfortunately, digitising the counting judgement task required animations, which was not supported by the digital platform we used. The results of the pilot study in Phase 1 identified, through descriptive statistics, that there were distinct tasks that did not function as intended (see example in Chapter 5.1.5.3). Furthermore, based on the subsequent analysis, most items were too easy for first-grade children in the final months of their first school year. Therefore, there was a need to adjust the difficulty level of the items and conduct a second pilot study with children in their final year of Norwegian early years education before the summer of 2020.

The pilot study in Phase 2 was conducted with 77 5- and 6-year-old children in Norwegian early years education in June 2020. The data collected in this study were used in Paper III to investigate what the digital assessment tool could tell us about Norwegian children's number sense. After the first days of data collection, the preliminary analysis identified a need to adjust the difficulty level: extending the number range in all categories and including symbolic representations in the arithmetic competence category. Therefore, I had to make some alterations to several items, making the total number of items in the study 71. Because of the update, the children were introduced to a maximum of 56 items. The specific findings from the second pilot study are described in Paper III. One of the main points for the further development of the assessment was that many of the items were still too easy for Norwegian first-grade children. Therefore, I had to adjust the difficulty level in the final version of the assessment. This was done by extending the number range and developing items to include the final three FoNS categories: representing numbers, estimation and number patterns.

To evaluate the quality of the final version of ENSA, I conducted a third pilot study in Phase 3. Qualitative observations indicated that some of the children accidentally found their way out of the assessment when the items were presented on a partial screen. Accordingly, I implemented full-screen functionality to make the assessment more accessible to the children. Additionally, I had to change the order in which some items were presented, ensuring that the first items in the assessment had the appropriate difficulty level according to most of the children's competence and avoiding too difficult tasks that might affect their motivation. After these changes were made, ENSA was ready for the main data collection. Before presenting the final assessment tool, I will introduce three aspects related to the design of the items.

5.1.5 Item design and considerations

During all three phases of the process illustrated in Figure 8, I designed the graphics for the chosen items based on the initial literature review, expert reviews and pilot studies using the program Inkscape (Inkscape Project, 2023). A programmer wrote the necessary JavaScript code for the items before they were made available for testing from a custom digital platform hosted by an NTNU-based web server. All items were adapted for the tablet format. Here, I will present considerations related to three central aspects of the item design: 1) graphics, 2) response alternatives, and 3) interactivity.

5.1.5.1 Graphics

The assessment tool was to be used individually by young children without needing instructions other than what was included in the tool. Therefore, it was necessary to find a way to provide instructions for the items and to send the children to the next item. For each item, I recorded the voice instructions, and the children were asked to press a blue arrow after providing an answer to an item to get to the next. The instructions were recorded and played to the children at the start of the assessment. In the pilot study in Phase 1, the children had to press a star in the upper left corner of the screen for instructions (see Figure 10a). The findings from this pilot study indicated that the children did not always press the star for instructions. Additionally, many children accidentally pressed the arrow at the bottom of the screen when they tried to provide answers to the items. It was crucial to include automatic instructions in the assessment tool to ensure that all children were given the same instructions. Moving the arrow at the top of the screen was also necessary. Furthermore, an agent was developed to symbolise the instruction button – making it more intuitive for the children that this was somewhere they could press to repeat the automatic instructions if they were not sure of the item content.

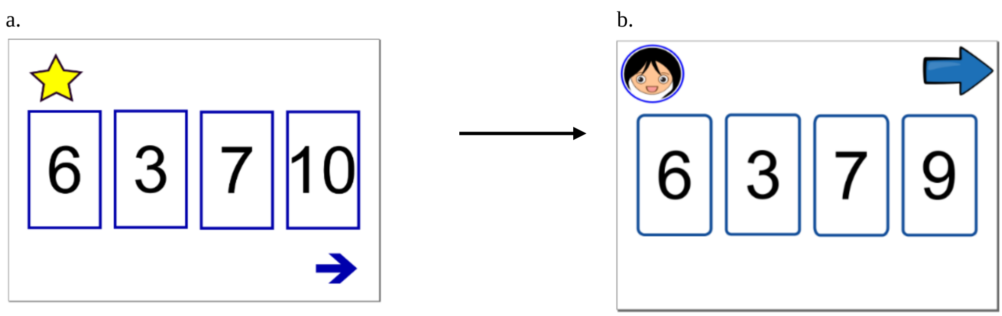


Figure 10. The development of an item from the pilot study in Phase 1 (a.) to the final version (b.). The star in the upper left corner on a. and the agent on b. were to be pressed for instructions. The pilot study indicated the need to change the arrow and the instruction button. In this item, the children would select the numeral that fit the instruction: 'select the number six'; it was possible to reselect before the arrow was pressed to get to the next item.

5.1.5.2 Response alternatives

In the pilot studies in Phases 1 and 2, I observed that some children used the response alternatives at the bottom of the screen in unintended ways when finding a solution to some of the assessment items. Because of the observations of the children's unintended use of response alternatives, it was necessary to examine how the ordering of response alternatives at the bottom of the screen might influence the children's responses to these items. Consequently, I designed items with different representations of response alternatives: ordered, unordered and no alternatives (see Figure 11). In the latter category of response alternatives, the correctness of the responses was evaluated based on the final position of the objects. In this way, we could evaluate how illustrations and symbols might influence the children's responses to the items. How these response alternatives might have affected the children's solution process is described in Papers 2 and 3. The different response alternatives and additional design elements related to the arithmetic competence category, and the effect these had on the item difficulty, are described further in Paper II.

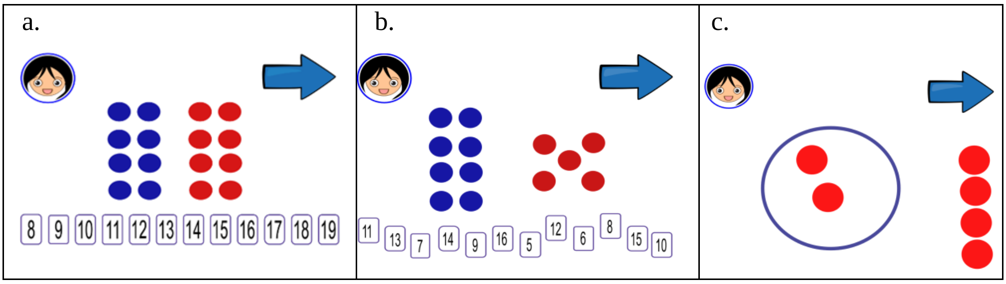


Figure 11. Three different representations of response alternatives: a. ordered response alternatives, b. unordered response alternatives and c. no response alternatives; the answer is evaluated according to the object's final position.

5.1.5.3 Interactivity

An important aim when investigating how a digital assessment tool can contribute to the formative assessment of first-grade children's number sense was to include interactivity in the assessment tool. A central hypothesis was that including interactivity could enable descriptions of steps taken towards finding a solution to an item. Therefore, each category was evaluated to identify possibilities for including interactivity. Simultaneously, finding a way to score the interactive items and evaluate the correctness of the responses was challenging. Consequently, there were two types of interactive items for which the final answer was scored by multiple choice or by evaluating the final position of the objects on the screen. Figure 12 illustrates two types of items from two different FoNS categories. Item a., from the number and quantity category, was scored by multiple choice and Item b., from the quantity discrimination category, was scored by evaluating the final position of the objects on the screen.

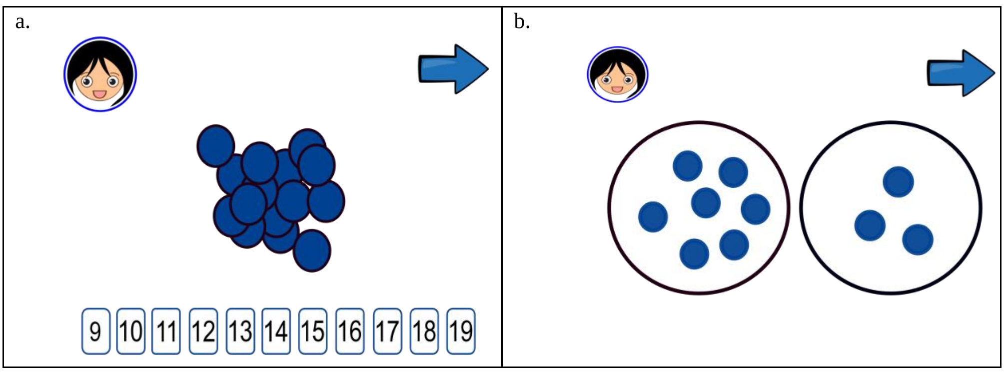


Figure 12. Two interactive items with two different ways of providing an answer. In item a., the children were to move the objects to count them and select the numeral according to the objects counted (multiple choice). In item b., the children were to move the objects so that there was the same number of objects in each circle (answer scored by final position).

One of the interactive items, in which the children's answers were scored by evaluating the final position of the objects on the screen, was one of the representing number items. This item, inspired by a task described by Andrews and Sayers (2015), as mentioned in Chapter 5.1.1.7, is illustrated in Figure 13a. In the pilot study in Phase 1, the first version of item a. was tested. Here, the children moved the objects on the orange line in different ways. An example of composing two and two was provided. The results showed that only two of the 32 children who participated in the pilot study understood or made any effort to compose the objects in different ways. Therefore, this task had to be developed further. Figure 13b. demonstrates the final version of the item included in ENSA, in which fixed groups of objects represented with dice enabled the children to find different groups of four.

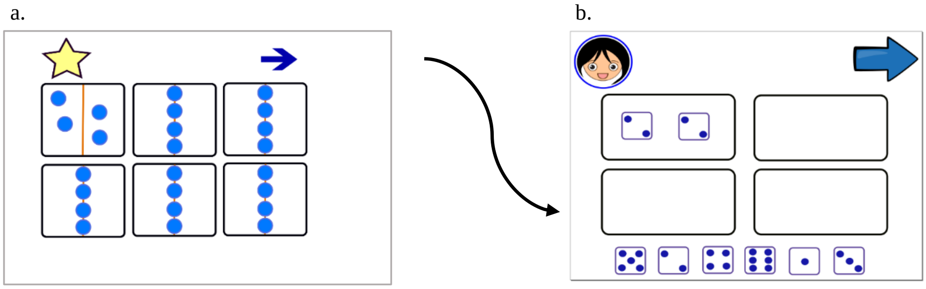


Figure 13. The development of an item in the representing number category called RN4, a. first version. b. final version (to the right). In a., the children were to compose four objects differently on each side of the line according to the first example. A further version of item b. was developed using familiar representations with fixed groups of objects (dice) to assist the children in grouping objects in different ways.

5.2 The final tool ENSA

Figure 14. The ENSA on an assessment continuum between static written assessments and more dynamic assessment methods, including observation of play and other activities.

One of the arguments in this thesis is that technology enables a combination of the advantages of written assessments and verbal assessment interviews. In Figure 14, I have placed the ENSA on an assessment continuum between written assessments and assessments through play or other activities. As described previously, a digital assessment, such as ENSA, has some practical advantages over written assessments regarding accessibility. Additionally, the challenges related to time at the other end of the continuum can be overcome using a digital assessment, such as ENSA, making the assessment easier to use. The possibility of interactivity places ENSA further towards assessment interviews and observation than traditional written assessments. At the same time, there are limitations related to digital assessment regarding context and descriptions of children's mathematical competence, which will be discussed in Chapter 7.

After developing the items for the final three FoNS categories and testing them with experts involved in the project, I decided which items to include in the final version of the assessment. The final overview of the 76 items included in ENSA is presented in Table 1. Some items had interactive (I) elements and recorded the children's movement of objects on the screen. Others were more regular items (R), which involved identifying a correct answer without any further process recorded. The interactive items are addressed qualitatively in Papers I and III and evaluated further in Paper IV.

Table 1. An overview of ENSA

Category	Description	N(R)	N(I)
Number identification (NI)	Recognise numerals and their meaning.	7	0
Systematic counting (SC)	Ordinality. Count to twenty and back from a random digit.	6	2
Number and quantity (NQ)	Cardinality. One-to-one correspondence between symbol and quantity.	6	4
Quantity discrimination (QD)	Compare quantities. Vocabulary: larger, smaller, more than, less than.	6	2
Arithmetic competence (AC)	Operate on small sets with addition or subtraction.	13	1
Estimation (ES)	Estimate the size of a set and position on a number line.	5	0
Subitising (SU)	Perceive quantity without counting. Perceptual and conceptual. Timed.	11	0
Representing number (RN)	Different representations of numbers and part–whole aspects.	4	1
Number patterns (NP)	Continue or complete a number sequence.	7	0

Note. An overview of the 76 items included in the tool with two types of items: Regular items (R), multiple-choice items asking the child to identify a certain number symbol or quantity, and interactive items (I), capturing the child's solution process.

5.3 Data collection

In both the second pilot study with the children in early years education and the main data collection with primary school children, the assessment was presented to each child on separate tablet computers using a custom digital platform. The participants were divided into groups of 2–6 children they were seated to avoid being disturbed by each other's screens or sounds. All the children were given the same instructions before they started the assessment.

Pre-recorded voice instructions were given for each item. For technical reasons, the assessment was presented in three separate units in the main data collection, with different FoNS categories and increasing difficulty. Each child could decide whether to continue to the next unit. Most children completed the first two units, some completed all three units and a few children completed only the first unit. No time limit was imposed for the items to avoid stress due to time pressure. The time on task was recorded for each item for research purposes. The children typically spent between 15 and 25 minutes on the assessment.

Qualitative data were gathered in both the pilot studies and the main data collection, taking field notes from each group session with the children. These field notes were further included in more extensive reports and later analysed for identifying aspects relevant for the development process. Furthermore, qualitative data from tasked-based interviews included in one of the master projects related to the thesis was included in Paper II (Schjølberg, 2021). In addition, qualitative data were collected through interviews with five teachers investigating teachers' use of assessment literacy and the instructional validity of ENSA.

5.4 Participants

In total, 498 individuals participated in various parts of this PhD thesis. The first pilot study included 32 participants in the first grade in a primary school in Malvik municipality. In the second pilot study, 77 children from five early childhood education institutions from Trondheim, Malvik and Stjørdal participated. The final pilot study included 16 participants from one primary school in Trondheim municipality. A total of 368 primary school children from eight schools in the three mentioned municipalities were included in the main data collection. Additionally, five primary school teachers participated in the qualitative study, looking into the instructional validity aspects of ENSA.

5.5 Analysis

In this thesis, I use quantitative and qualitative data collection methods, and these types of data need to be analysed differently. Hence, I apply the dichotomous Rasch model to analyse the children's responses to ENSA and perform a thematic analysis of the data from the teacher interviews.

5.5.1 Rasch analysis

The data from the main data collection and the pilot studies were analysed using Winsteps software (Linacre, 2022). Winsteps implements the joint maximum likelihood estimation algorithm to estimate the parameters of the Rasch model. The Rasch model was further described in Chapter 4.5. To evaluate the content, substantive and structural aspects of the validity of ENSA (further described in the next chapter), I performed an analysis of person and item fit in addition to an analysis of the dimensionality of the data.

The fit of the person and item measures can be evaluated to determine the overall usefulness of the scale concerning the consistency of the underlying structure of the data according to the expectations of the Rasch model. Fit measures are examined by evaluating the infit and outfit mean-square (MNSQ) statistics. Values between 0.5 and 1.5 for MNSQ statistics are considered productive for measurement, while values less than 0.5 are less productive but not degrading (Wright & Linacre, 1994). Therefore, I concentrated on MNSQ values greater than 1.5 to diagnose the misfit.

Unidimensionality is a precondition of Rasch measurement and can be evaluated by an interpretation of the principal component factorial analysis of the standardised residuals to examine whether one dimension effectively explains the variance in the data or whether there are secondary dimensions or contrasts (Smith, 2004). On a unidimensional scale, the expected eigenvalues explained by the first contrast are around 2.0 (Linacre, 2022). In addition, the residual values should follow a normal distribution and represent random noise, and person responses that fit the Rasch model should be independent of each other or uncorrelated (Schumacker, 2004, p. 247).

5.5.2 Thematic analysis

The interviews were analysed by applying a thematic analysis following the stages of the thematic analysis process (Braun & Clarke, 2006). I have described the theoretical perspectives related to the analysis in Chapter 4.6.

In the first stage of the analysis, I transcribed the data in the original Norwegian language before discussing the translated data transcripts with two of my supervisors. Second, initial codes were developed while looking for duplicates. As a third step, I started looking for related themes and grouping codes. The fourth step involved sharing the initial codes and related themes with one of my colleagues – confirming that the themes were present in the data. The themes were further shared and reviewed in discussions with one of my supervisors in the fifth step of the analysis. Finally, the sixth step involved selecting relevant quotations to formulate the analysis's story. In this way, I tried to examine the social, temporal and physical boundaries in the teachers' interpretations of the assessment results. As a final step, the statements from the teachers that formed the natural unit and the themes and interpretations were translated into English.

5.6 Validity

Qualitative and quantitative data require different kinds of analysis. The methods of data collection are also tied to conflicting paradigms and, consequently, different views on aspects of validity. In this section, I will describe validity aspects related to the quantitative approaches to investigating the validity of ENSA. Furthermore, I will investigate how the qualitative approach to investigating the teacher's perspective on ENSA necessitates a different perspective on validity tied to trustworthiness and authenticity.

5.6.1 A quantitative perspective on validity

In this thesis, I investigate the validity of ENSA by applying specific descriptions of validity in a Rasch measurement and test development context (Wolfe & Smith, 2007b) using different data collection methods. The validity concept refers to specific definitions and distinct courses of action depending on the chosen research methodology. In this thesis, I have considered seven aspects of validity based on the applied validity approach and addressed the content, substantive, structural, generalisability, external, consequential and interpretability validity aspects of the assessment tool through the five included papers (see

Figure 15). The responsiveness aspect of validity, related to detecting change after the intervention, is the only validity aspect from the applied framework that is not addressed in this thesis. On another level, I have also considered the validity of my project, especially the validity of the final study with the five teachers.

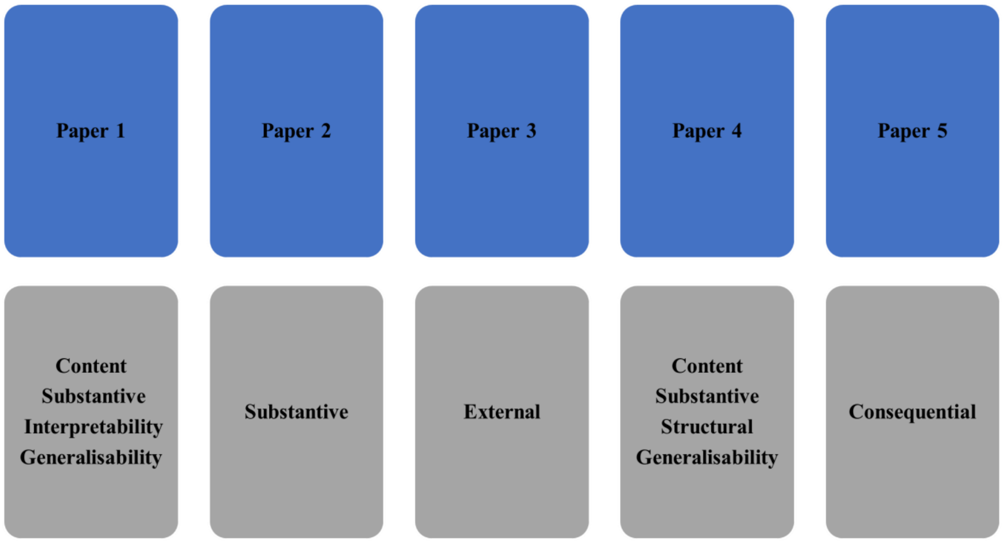


Figure 15. An overview of how the five papers addresses the seven aspects of validity (content, substantive, structural, generalisability, external, consequential and interpretability).

In the first paper, I provide details regarding evidence of content validity, such as item development, instrument purpose, test specifications, expert review and item technical quality. I address the substantive aspect of validity related to the theoretical rationales and further evaluate how our data fit the Rasch model by looking at validity evidence, such as person fit and item difficulty hierarchy. In this paper, we also describe that the process items, later called interactive items, can provide teachers with more information regarding children’s solution process and therefore add to the evidence of the interpretability aspect of ENSA. In the fourth paper, I describe how interactive items may provide evidence for the substantive validity of the assessment tool and inform about the process of children when solving digital assessment items. Moreover, interactive items may provide evidence for assessment validity’s interpretability and substantive aspects.

The second paper investigates the substantive aspect of validity and focuses specifically on the item difficulty hierarchy of arithmetical competence items in ENSA. Validity is not mentioned in the manuscript, as I focus on describing how the various design elements and problem types affect the difficulty of the items.

In the third paper, although it is not mentioned specifically in the manuscript, the external aspect of validity is addressed, as I, with this contribution, have investigated the external validity of ENSA in two different groups of children in early years education and the first grade of primary school.

In the fourth paper, I provide validity evidence relevant to the content, substantive and structural aspects of the validity of ENSA, focusing especially on the validity evidence of the interactive items. Regarding content validity, I evaluated the relevance and representativeness of ENSA and analysed the technical quality of the items. Substantive validity is addressed by considering children's interpretations of the various items (person fit and item difficulty hierarchy). Looking at the structural aspect of validity, I evaluate the trustworthiness of the scoring structure by a dimensionality analysis of the assessment tool. Additionally, the generalisability aspect is addressed in this paper by presenting the reliability measures of ENSA. Reliability measures are also presented in Paper I.

Evaluating the consequential validity aspect of ENSA in the final paper, I had to look at other sources of evidence for the consequential validity of the tool, as Rasch analysis of the data could not provide evidence of consequential validity. Applying an additional framework for instructional validity, I, therefore, interviewed five teachers to enable the analysis of their descriptions of how they would use the assessment tool to enhance their children's further learning and what could affect their interpretations. This paper also addresses the interpretability aspect of validity, as I look at the possibility of presenting children's results to teachers using person-item maps. The quality of the data in this paper is further addressed in the next section.

5.6.2 The quality of the qualitative data

Addressing the quality of qualitative data is something different than looking at the validity of quantitative data; therefore, in line with Lincoln and Guba (1985), I have considered the trustworthiness and authenticity of the qualitative aspects of the thesis.

A consideration of the trustworthiness of the qualitative data includes the four criteria of credibility, transferability, dependability and confirmability (Lincoln & Guba, 1985). Here, credibility refers to the research practice and how the data reflect the actual meanings of the participants. While conducting the interviews with the teachers, I was conscious of any ambiguity and repeated the participants' statements to get their confirmation in terms of communicating their real meanings. In addition, I have focused on criteria transferability by providing detailed descriptions of the research context to enable the comparison of the results to other contexts. I have also kept a record of all process phases, providing details regarding transcripts and analyses to ensure dependability. The interview guide used in the interviews, excerpts from the transcripts and descriptions of the analysis process are all presented to ensure transparency so that others can investigate my research process. Concerning ensuring transparency and the final criteria confirmability, I have also discussed the data and resulting codes from the thematic analysis with my colleagues to obtain a richer view of the data from different perspectives. By obtaining different perspectives on the qualitative data, I have tried to ensure objectivity. While no interpretation can be free from subjectivity, I have ensured different perspectives on the materials to decrease the influence my personal opinions could have on the interpretations of the teachers' perspectives.

I have also considered the authenticity of the qualitative data concerning fairness. I interviewed five teachers from different schools to capture different points of view on children's results from ENSA. This also aligns with my pragmatic paradigm, as described in Chapter 4. Authenticity comprises ontological, educative, catalytic and tactical aspects (Lincoln & Guba, 1985). These aspects relate to the participants' understanding of their context and other contexts and how participation in research might engage the participants in action to change their circumstances. While I cannot say anything about how the teacher's participation in the interviews might have affected their understanding and engagement of the assessment context in retrospect, their engagement in the discussed topics was evident during the interviews. Furthermore, the aim is that the results from the thesis can engage other teachers in reflecting on and developing their assessment literacy.

5.7 Ethical considerations

Throughout my work with this thesis, I have been conscious of my role as a researcher. My background as a primary school teacher was of great value when connecting with children and teachers and seeing how my visit to the school affected their daily routines.

From recruiting participants for the studies, carrying out the assessment with children and conducting interviews with the teachers, I have focused on the participants' experiences when participating in the study. Before each session with the children, we had an informal chat in which I explained what we would do and why. Moreover, I talked to the children about how carrying out the assessment was voluntary and how they could finish it at any time. It was also important for me that the children, both in early years education and primary school, were presented with the assessment tool in known everyday surroundings, with familiar adults available if they needed assistance. Additionally, all the teachers who participated in the interviews were given the required information before the interview started, gave their informed consent and were free to withdraw from the project at any time they wanted. To avoid any harm to the participants caused by the interviews in the form of stress, the participants chose the time and place of the interviews. After the participants had consented, the interviews were recorded and transcribed to ensure anonymity.

The work related to the thesis has been approved by the Norwegian Centre for Research Data (NSD), and the necessary guidelines related to depersonalisation and parental consent have been followed (see Appendix B and C). All the data were stored according to the data management plan approved by the NSD. No personal identifiable information is presented in the publications resulting from the project. Still, to provide detailed descriptions of my research context, I have included information regarding the municipalities to which the various schools belonged. All three municipalities included in the study are quite large, and including this information in the papers did not affect the anonymity of the participants.

5.8 Chapter summary

The development of ENSA has been a major part of the thesis; a detailed description of the development process of ENSA has therefore taken up much of the space in this chapter. When the specifics of the development process are not described in any of the papers included in the thesis, I needed to include more details on the assessment development in the methods chapter to provide detailed descriptions of my work and ensure transparency. Further in this chapter, I have presented details regarding data collection, project participants, analysis and validity to describe my ways of investigating ENSA and my role as a researcher. This role is central to my final reflections on the ethical considerations connected to the thesis.

6 Summary of the five papers

In this chapter, I summarise the five papers included in the thesis. From different perspectives, all the papers investigate the development, validation and use of ENSA. I provide a brief introduction to the background, research questions and the methods applied in each of the papers before I present the results, conclusions, and further implications.

6. 1 Paper I: Developing a formative, digital tool to assess children's number sense when starting school

Paper I was published in the proceedings of NORMA20 (Nortvedt et al., 2022). In this paper, we investigate the opportunities in technology for aiding teachers in their formative assessment practices and address the need to "develop a diagnostic tool for teachers to assess individual grade one children's FoNS-related understanding" (Sayers et al., 2016, p. 389). Operationalising number sense with the FoNS model (Andrews & Sayers, 2015) and focusing on the content, substantive, and interpretability aspect of validity (Wolfe & Smith, 2007a), we pose the following research question: How do digital tasks support aspects of validity in the assessment of first graders' number sense?

Using preliminary data from the main data collection, we present data from 101 first-grade children who carried out the assessment, later to be named ENSA. The development of the assessment tool, related pilot studies and expert review were presented as part of the results, in addition to an overview of the items included in the tool. Furthermore, we presented two types of items in the tool: skill-based items and process items (later referred to as regular/traditional items and interactive items). We displayed the development of one process item and provided examples of how two children solved this task. We discussed how the information provided from these examples could enable support for the substantive validity of the assessment, in addition to being valuable for screening, individual follow-up, and teaching purposes.

The dichotomous Rasch model (Rasch, 1960) was used to analyse the children's responses using the Winsteps software (Linacre, 2017). We evaluated the quality of the assessment tool by addressing the technical aspects of construct validity. The person reliability value of the 76 items analysed in this paper was 0.88 (corresponding to a Cronbach's alpha of 0.91). The items were well targeted to the group of first-grade children and were slightly easy. For this assessment tool, most items fit well with the Rasch model. One of the representing number items was presented as an example of misfitting items, and the reasons for this are discussed. The interpretability aspect of validity was also supported by FoNS categories with different ranges of difficulty; difficulty correlated with the numerical value of the answer, indicating what number range the children could confidently process, although this correlation was not present in all categories.

We conclude that a digital assessment tool has the potential to provide teachers with information about their students' number sense and supplement formative assessment validly and reliably by combining advantages from other assessment formats. Our results show that the application of interactive items can reveal more of the children's solution process, adding to the interpretability and substantive aspects of assessment validity. The content and substantive aspects shed light on in this paper are further investigated in Paper IV. While investigating the development of ENSA, I saw that I needed to look more specifically at how the different design elements used in the various tasks in the assessment affected the children's responses; this was the background for writing Paper II.

6.2 Paper Ⅱ: Factors that influence the difficulty level of digital arithmetic assessment items for first-grade children

Published in the proceedings of CERME12 (Hodgen et al., 2022), Paper II follow up on the need to investigate the affordances and constraints by applying digital technology to mathematical assessment items. We look specifically at the items designed to measure arithmetic competence as a component of the FoNS model (Andrews & Sayers, 2015) to answer the question: Which design elements influence the level of difficulty of digital arithmetic assessment items, and how do the design elements influence the strategies that first-grade children use to solve these items?

The arithmetic competence items in ENSA were designed to enable the investigation of how problem type (Carpenter & Moser, 1984), and magnitude of the answer, in addition to various design elements such as the use of symbols, pictures and response buttons with ordered or unordered numerals, affect the difficulty level of the items. Four types of arithmetic problems were included in the design: Change and combine (sum items) and compare and equalise (difference items). A total of 302 first-grade children solved fifteen arithmetic items as part of the main data collection presented in the thesis. Groups of six to eight children carried out the assessment on separate tablet computers. Qualitative data from individual interviews conducted with 19 first-grade children solving the arithmetic items were collected independently as a part of a master's degree project (Schjølberg, 2021). The goal of the interviews included in the master's project was to get an overview of the children's strategies. One of the strategies applied by one of the students who participated in the master's project is included in this paper. The quantitative data was analysed using Rasch measurement, and the qualitative data was analysed using thematic analysis.

We found that the strongest determinant of item difficulty was the problem type. The four difference items were also the four most difficult arithmetic items. Within the difference category, both compare items had higher difficulty than the two equalise items. The compare items were more difficult despite involving small numbers, while the equalise items involved large numbers and came with more complex voice instructions. We discuss the reasons for these differences in difficulty and indicate that simplified instructions in word problems may lead to more misunderstandings and reduce the child's possibilities for modelling the situation. The use of illustrations and the children's previous experiences are also presented as reasons for the difference in difficulty. The second strongest determinant for item difficulty was the numerical value of the item's answer: We found that difficulty was strongly correlated with the magnitude of the answer of an item. Whether the problem was presented in a symbolic or pictorial form did not affect item difficulty. Additionally, our data indicated that ordered response buttons allow children who have not mastered large numerals to use these buttons as a number sequence that helps them solve the problem.

These results underline the importance of carefully investigating the various design elements, the magnitude of answers, number representations, and ways to report answers when developing digital assessment items.

6. 3 Paper Ⅲ： Five-year-old children’s number sense in early childhood education using digital tasks

In this paper, published Nordic early childhood educational research (Saksvik-Raanes et al., 2023), we investigate what we can learn about children’s number sense in early years education from digital tasks. There are many descriptions of the valuable number sense experiences children acquire from an early age. However, the activities included in early years education often underestimate the children’s knowledge. We aim to investigate a new method for describing young children’s number sense using a digital assessment tool and ask the following research question: What can we learn about 5-year-old children’s number sense using a digital assessment tool for early number sense? In this way, we study if the tool can contribute to providing an overview of young children’s number sense in a more time-efficient way than other forms of assessments.

From an early number sense perspective, we present the FoNS model that represents the number related activities children learn at the start of formal schooling and connect the categories of the FoNS model to the curriculum of Norwegian primary and early years education. Further, we present the current knowledge about children’s number sense development in Norway and the international context.

The data collection was conducted with 77 five-year-old children from five early childhood education institutions during the second phase of the development of ENSA comprising 71 items from five FoNS categories: number identification, systematic counting, number and quantity, quantity discrimination, and arithmetic competence, in addition to subitising. The items were presented to the children with tablet computers, and the data were further scored dichotomously and analysed using Rasch measurement. We also include qualitative data from observation notes and reports made during the data collection. These notes and reports were analysed to identify children’s strategies and perspectives on the different items in the tool.

The results revealed that many of the children in our study knew more about numbers than expected in the curriculum for Norwegian early years education and the first two years of primary school; also, there were large individual differences between children already before the start of formal education. Qualitative observations indicated that quantitative data collection methods could not describe some aspects of children’s interactions with digital tasks. This is illustrated with an example of how the participant “Petter” uses the representations available to find the numeral representing the counted quantity without knowing the symbolic representation for this number. Compared to what is known about Norwegian children’s development of number sense in early childhood education, the results of this study are surprising. We discuss what the results could mean for facilitating children’s number sense experiences in early childhood education.

6.4 Paper IV: Measuring children's early number sense with digital tasks: Evaluating the validity evidence of a digital assessment tool with interactive items

In this paper, submitted to ESM, we evaluate the validity evidence of ENSA to investigate the potential of digital assessment tools and the challenges that come with this assessment format. Digital assessment tools can overcome some of the challenges in traditional assessment formats and look deeper into children's knowledge; simultaneously, there are challenges connected to translating assessment items from other formats and confounding technological proficiency with mathematical proficiency. To answer these challenges, we looked further into the content and substantive aspects of validity described in Paper I, in addition to structural validity. We investigated if ENSA could tell us more about assessing children's number sense in a way that displays validity evidence while including interactive items that record children's solution processes. In the paper, we address three research questions: 1) What is the evidence for the validity of ENSA–a digital tool for assessing early number sense? 2) Can the interactive items be considered part of the same construct as the regular multiple-choice items? 3)What is the added value of integrating interactive items into the assessment?

The responses of the 368 first-grade children to the 76 items available on tablet computers were scored dichotomously and analysed using Rasch measurement. We evaluated the children's responses to the regular and interactive items in ENSA by addressing the content, substantive, and structural aspects of validity. We also included qualitative descriptions of two children who responded to one of the interactive items.

Through the evaluation of ENSA, we provided evidence that digital assessments of early number sense can be performed with a high degree of content, substantive and structural validity. The dimensionality analysis indicated that technology was not a distracting factor or a second dimension in the assessment process in our sample of children. We therefore consider the interactive items to measure the same number-sense construct as the regular multiple-choice items. The interactive items provided additional value to the assessment in at least two ways: i) they improve the accessibility of the assessment by allowing children to express their knowledge of non-symbolic mathematical procedures—such as matching, structuring and counting—even in the absence of knowledge of mathematical symbols and numerals, and ii) they allow teachers or researchers a record of the steps taken towards a solution—rather than simply the final response—for qualitative analysis.

The paper also contributes to the field of digital assessment by showing that interactive items can be validly interleaved with more traditional items, which is an important contribution to developing digital assessments for primary mathematics education with strong validity evidence.

6.5 Paper V: Factors influencing teachers' interpretations of assessment results

In the final paper in the thesis, we investigate the opportunities for teachers to use ENSA for formative assessment and factors that might influence this use by answering the following research question: What factors influence teachers' interpretations when reflecting on the assessment of children's number sense?

Previous research has identified how teachers experience challenges in implementing formative assessment connected to knowledge of the subject, the perceived value of test results, and teachers' access to support materials. These challenges are all connected to the situated knowledge of teachers' assessment literacy. How teachers interpret and use assessment in mathematics education is poorly understood, and further insights are needed to design diagnostic assessment tests to support teachers' use of assessments.

Five first-grade teachers were presented with children's results on ENSA intended to inform their instructional decisions. In semi-structured interviews with the five teachers, the results were presented through a modified version of a KIDMAP from the previous Rasch analysis of the data from the main data collection, indicating surprising answers from the children based on their overall number sense knowledge. Further, a thematic analysis was conducted to identify the factors that influenced the teachers' interpretation of the assessments.

Findings indicate that the instructional materials chosen by the school limited and shaped teachers' interpretations of children's assessment results. Teaching objectives linked to the start of formal schooling reduced the formative value of assessments for teachers. External factors like parent expectations and assessment context limited teachers' interpretations of the evidence. Furthermore, teachers requested a larger focus on the potential of all children and a need for guidance for the formative use of assessments. Together, these factors form a challenge to initiatives to enhance formative assessment by teachers. We highlight the need to investigate the instructional validity and factors influencing teachers' use of such assessment instruments.

7 Discussion and conclusion

In this thesis, I set out to investigate how a specific digital resource can contribute to the formative assessment of children's number sense at the beginning of the first grade. I could not find an assessment tool that was 1) in a digital format, 2) adapted to the Norwegian context and 5–6-year-old children, and 3) focused specifically on describing children's number sense while displaying evidence of validity. Therefore, I developed ENSA as a research tool investigating how such tools can contribute to the formative assessment of children's number sense.

I found that the specific digital assessment tool can contribute to the assessment of number sense by offering new possibilities in interactive items for describing further aspects of children's number sense without limitations related to time and format. The papers also shed light on how applying a specific digital assessment tool to formatively assess children's number sense comes with challenges. Newly introduced design elements in the digital assessment context might be used in different and unintended ways by the children and can therefore affect the validity of the assessment. Additionally, teachers' application of their assessment literacy in formative assessment practice is situated in a wider social context and is, therefore, a complicated process. It is important to consider this complexity when designing assessment tools and evaluating the evidence of validity.

7.1 Summary of findings and answers to research questions

The five papers in this thesis shed light on how technology can usefully supplement the assessment of young children's number sense by describing ENSA's development and validity analysis.

7.1.1 How do digital tasks support validity aspects in assessing first graders' number sense?

A description of the first validation of ENSA is presented in Paper I by addressing three validity aspects (Saksvik-Raanes & Solstad, 2021). We found that the digital tasks supported aspects of the content, interpretability and substantive aspects of the validity of the assessment by offering tasks of high technical quality. Through addressing the interpretability aspect of validity, we found that the tasks provide a strong evidence base with the potential to inform teachers about children's number sense by describing different ranges of difficulty in different FoNS categories, variation between categories and how much numerical value matters for difficulty. Moreover, we found that digital tools have the potential to adapt tasks previously reserved for more dynamic assessment formats, such as individual interviews. Finally, investigating the substantive aspect of validity, we found that digital tasks can provide information about the process behind children's answers to assessment items for screening, individual follow-up and teaching purposes.

7.1.2 Which design elements influence the difficulty level of digital arithmetic assessment items, and how do the design elements influence the strategies that first-grade children use to solve these items?

The specific design elements included in ENSA are investigated in Paper II (Saksvik-Raanes & Solstad, 2022). Specific design elements such as problem type, representations, numerical values, and differently ordered response buttons influence digital items' difficulty levels. Additionally, our analysis indicates that one of these elements, the ordered response buttons, affects how children find answers to digital items. More specifically, we found that problem type was the strongest determinant of item difficulty: difference items were more difficult than sum items, and compare items were more difficult than equalise items. The second strongest determinant of item difficulty was the numerical value of the item's answer. Whether the problem was presented in a symbolic or pictorial form did not affect item difficulty.

Finally, we found that the ordering of response buttons may influence how children find the correct answer for an item. Our qualitative data indicate that ordered response buttons allow children who have not yet acquired mastery of large numerals to use these buttons as a number sequence that helps them solve the problem.

7.1.3 What can a digital assessment tool tell us about five-year-old children's early number sense?

The digital assessment tool provided insights into three aspects of the early number sense of five-year-old Norwegian children, as presented in Paper III (Saksvik-Raanes et al., 2023). First, we confirmed findings from previous research regarding the large spread in what children know at this age, from children with limited understanding regarding symbols and numbers larger than five to children displaying competence in solving two-digit arithmetic problems. Furthermore, we found that the children as a group displayed a competence that exceeded the descriptions in the national curriculum and previous research in the Norwegian context. Additionally, our qualitative data showed how children could use the representations available as a number sequence to solve problems beyond their knowledge of large numerals. While the study is limited to the applied assessment tool, Norwegian context and the number of participants, our results show that new methods for investigating children's mathematical competence may provide further insights into children's early number sense development.

7.1.4 Three related questions on the interactivity and validity of ENSA

Further validity evidence of ENSA, focusing on interactive items, is presented in Paper IV. First, we addressed the following question: What is the evidence for the validity of ENSA – a digital tool for assessing early number sense? Here, we found that the evaluation of ENSA provided evidence that digital assessment of early number sense can be performed with a high degree of content, substantive and structural validity. Furthermore, we investigated whether the interactive items could be considered part of the same construct as the regular multiple-choice items. The dimensionality analysis indicated that technology was not a distracting factor or a second dimension of the assessment process for the participants in our study. Therefore, we considered the interactive items to measure the same number sense construct as the regular multiple-choice items. Finally, we considered the added value of including interactive items in the assessment. The interactive items provided additional value to the assessment by improving the accessibility of the assessment, allowing children to express their knowledge of non-symbolic mathematical procedures through matching, structuring and counting. Thus, including interactive items allowed the children to communicate their understanding without knowing mathematical symbols and numerals. Additionally, the interactive items allow teachers or researchers to record the steps taken towards a solution rather than simply presenting the final response and using this recording for qualitative analysis.

7.1.5 What factors influence teachers' interpretations when reflecting on the assessment of children's number sense?

Teachers' interpretations and reflections on children's assessment results are complex processes affected by teachers' assessment literacy situated in a social context. In Paper V, we identify central factors that influence teachers' interpretations of children's number sense assessment results: instructional materials, perceptions of equality in school, construct irrelevant factors, parental expectations and a need for guidance towards the formative focus of assessments. In this regard, a misalignment between the instructional materials and the curriculum poses a serious threat to assessment validity and the decisions made based on the assessment process. Furthermore, the teachers had mixed feelings about their knowledge and practice when meeting children at different starting points in first grade in a Norwegian school system focusing on inclusion and equity. Additionally, the teachers described different external factors as reasons behind a child's answer, causing a threat to instructional validity related to the administration of assessments. Finally, the teachers requested more focus on children's potential, solution processes, and guidance to facilitate further learning. A lack of insight into the background of children's responses threatens the instructional validity of assessments related to scoring. As discussed in previous papers, the digital format can ameliorate this issue by introducing new opportunities for teachers to gain insights into the background of children's responses and to use this information in their formative assessment practice.

7.2 Conceptual, methodological, and empirical contributions

The findings from this thesis bring theoretical, methodological and empirical contributions to research on assessment in mathematics education. Operationalising number sense in an assessment tool using the FoNS model within a wider perspective of mathematical proficiency brings new opportunities for gaining insights into the number competence of first-grade children. This thesis contributes to methodological advances by introducing a method to describe children's number sense with the help of technology. Furthermore, using these theoretical and methodological insights provides empirical contributions by describing new aspects of children's number competence.

7.2.1 Conceptual

To my knowledge, ENSA is the first assessment tool developed based on the FoNS model. Applying ENSA, or other tools based on specific number sense models, to investigate children's number sense can contribute to making a messy construct more useful for teachers. By operationalising number sense through eight categories and presenting children's assessment results from each category, ENSA can help teachers develop targeted and adapted teaching. Thus, ENSA can become a tool to assist teachers in their formative assessment practices. Moreover, applying a culturally adapted model, such as the FoNS model in the Scandinavian context, contributes to making the construct relatable for teachers. The FoNS model was originally developed in a Scandinavian research context to identify children's learning opportunities in different cultural settings (Andrews et al., 2015; Andrews & Sayers, 2015). Therefore, describing children's learning opportunities based on models that relate to the specific cultural context where the assessment is to be used can contribute to making the number sense construct more relatable.

The thesis also contributes to viewing number sense from a wider perspective, comprising central dimensions of FoNS (Andrews & Sayers, 2015) and mathematical proficiency (Kilpatrick, 2014) to number competence. Children's number competence is in this thesis more than foundational or early number sense and relates to mature and applied number sense (Andrews & Sayers, 2015; Whitacre et al., 2020) – using number sense flexibly in one's daily life. A connection between models of number sense and frameworks of mathematical proficiency can clear up some confusion connected to the various descriptions of what constitutes children's number sense. Additionally, the connection between these perspectives can inform further research on children's number sense development, the central skills included in the concept and the wider mathematical competence necessary to apply these skills in flexible ways.

The conceptualisation of children's number sense through the FoNS model and number competence in a wider perspective can clarify a messy construct for teachers and enable further investigation of children's number competence. Learning mathematics is a complex process, and the methodological contributions of this thesis present new methods for assessing this knowledge to shed light on children's mathematical competence.

7.2.2 Methodological

This thesis introduces a digital tool applied to investigate young children's number sense without limitations related to assessment format or time. Applying technology in the assessment of young children's number sense enables the description of the competence of large groups of children while displaying evidence of validity (Saksvik-Raanes & Solstad, 2021). In this way, we can describe new aspects of children's development of number sense unhindered by limitations regarding children's reading and writing skills or the time demands of individual assessment interviews.

The findings presented in this thesis indicate that digital assessment tools, such as ENSA, can provide information about children's abilities connected to the measured number sense categories, as well as how children use their number sense to sort out objects in the solution process and apply representations to provide answers to tasks (Saksvik-Raanes & Solstad, 2021). Additionally, analysis of the dimensionality of the measured number sense construct indicates that interactive items measure the same construct as traditional assessment items (Saksvik-Raanes & Solstad, 2023). In this way, the results point to the potential of technology to describe the complex mathematical competence children apply when solving assessment tasks. This aligns with previous research pointing to this potential (Drijvers, 2018; Ginsburg, 2016; van den Heuvel-Panhuizen et al., 2011).

7.2.3 Empirical

By operationalising the early number sense of Norwegian children, ENSA provides a method for describing the number sense of children before they can read or write. The findings indicate that the children might have a more developed number sense than described in previous research and the descriptions in Norwegian early childhood education and primary school curriculum (Saksvik-Raanes et al., 2023). These findings contribute to greater insight into children's development of number sense in the Norwegian context. Looking forward, there is a need to investigate whether the findings from this study hold up in a larger sample. If that were the case, this would have major implications for teaching and learning mathematics in Norwegian first grade, as discussed in Chapter 7.4.

7.3 Limitations

The methods applied in this thesis also set some limitations in investigating how ENSA can contribute to the formative assessment of children's number sense. Applying different methods could enable the investigation of other perspectives of the tool in use. Based on my descriptions in Chapter 3, looking at the assessment of number sense as a process and a product, there are some aspects that my applied methodology cannot shed light on. Looking at human reality as different constructions will not enable descriptions of the social practice that teachers, children and various tools are a part of. I can only describe the children's number sense at one point in time, within specific conditions and how five teachers interpret these results. There is a need for different lenses to describe how teachers use formative information in their teaching practice. In my data, there is a gap between what teachers reflect on and their classroom practices. The results of this study could have been strengthened by including observations as a third source of generating data for triangulation purposes.

To preserve the anonymity of a small sample of children from a geographically restricted population, I did not collect any demographic information involving age, gender, or social background in this thesis. However, developing ENSA opens possibilities for asking new sets of questions. Applied to a wide population, collecting demographic information could enable investigating issues related to geography, socioeconomic status, and equity. Previous research has, for instance, looked into how number-related activities in the home environment affect children's number sense (Phillips & Crowell, 1994). In a Rasch measurement context, demographic information could enable the evaluation of the differential item functioning – investigating how the items in ENSA functioned between different groups and evaluating whether there were any biases related to specific assessment items. Such investigations can contribute to developing assessment items that enable all children to communicate their knowledge of numbers independent of their background. Furthermore, investigating how different items function between different groups of children can also permit descriptions of how to support children from different backgrounds best to develop their number sense.

In this thesis, insights into children's number sense are limited to the dimensions of the FoNS model. There are possibilities and limitations to applying the FoNS model in this investigation. As discussed in Chapter 7. 2. 1, looking at number sense through eight different categories may be useful for teachers as a conceptualisation for describing children's learning opportunities. At the same time, the measurement of these categories, in addition to subitising, resulted in many items and further challenges related to dimensionality (Saksvik-Raanes & Solstad, 2023). Additionally, the FoNS framework and the operationalisation of the categories through the items in ENSA rely heavily on symbolic representations. Other models, focusing on children's fingers and embodied representations, represent a further potential for investigating children's number sense, as discussed in the next chapter.

7.4 Implications and further work

Through this thesis, I have investigated interactive items and shed light on the potential of technology for describing children's processes of solving assessment items. Additionally, I have described how such items can be used to provide teachers with information regarding the competence of individual children. In this way, information from digital assessments can be used for formative assessment practices and to compare different solution processes for teaching purposes and research. The described potential in these interactive items calls for further research to investigate the opportunities for developing assessment items that can tell us more about young children's number competence.

Through qualitative observations included in Paper II and Paper III, we saw that there were aspects of the children's solution process that ENSA did not capture. Some items were not possible to develop due to limitations on the applied platform when including animations (see Chapter 5. 1. 4). Furthermore, many of the items in ENSA were dependent on the children's knowledge of symbolic representations of numbers. In further investigation of children's number sense using digital tools, there are possibilities in applying other frameworks to include some of the aspects ENSA could not capture. Recently, there has been an increased interest in children's use of finger representations. Using rhythm as the main unit of analysis, Sinclair et al. (2016) looked at children's engagement and joint activity while exploring the TouchCounts application (Sinclair & Jackiw, 2014). Framed by the theory of semiotic mediation, Baccaglini-Frank et al. (2020) described a schematic organisation of number sense abilities focusing on ordinality, finger gnosis, finger control and finger representations of numbers. Others have also emphasised the importance of fingers and gestures in children's development of early number sense in the frame of embodied cognition (Tucker & Johnson, 2022). Research on applications such as TouchCounts (Sinclair & Jackiw, 2014) and Fingu (Holgersson et al., 2016) has shown a vast potential in multi-touch technology to investigate qualitative aspects of children's number sense. By applying models that focus more on children's embodied representations, it is possible to develop assessment tools to capture children's use of physical representations and rhythm in their interactions with mathematical tasks. Combining multi-touch technology with frameworks related to innate number sense or children's spontaneous focus on numerosity (Hannula-Sormunen et al., 2015) provides new opportunities to describe young children's number sense development without relying on symbolic representations.

By investigating how different design elements and various representations affect assessment difficulty in Paper II, in addition to examining the validity evidence of interactive items in Paper IV, I have addressed issues related to how digital assessments might assess mathematical or technological skills and how technology can limit or enhance the means of expressing oneself mathematically (Drijvers, 2018). This thesis indicates that children's use of the available representations when solving mathematical problems may threaten assessment validity. In developing digital assessment tools that include new elements, contexts and interactivity (Ginsburg & Pappas, 2016), it is important to investigate how these factors influence children's interactions with the tool. Further research must address how introducing new representations and designs in a digital context may influence how children find answers to assessment tasks and ultimately affect the validity of assessments.

There is wide agreement that children have a well-developed number sense before the start of formal education, and the findings from this thesis support this view. We examined the possibilities of describing young children's knowledge with a new assessment method and found that Norwegian 5-year-old children might know more about numbers than one usually expects when children start school in Norway. These results call for further research to investigate whether Norwegian children are met with activities adapted to their number sense in early years education and primary school. In this way, the findings from this thesis can contribute to equity in the mathematics classroom - ensuring that all children are met with equal opportunities to develop their number sense based on their previous knowledge.

Findings from Paper V indicate that the teachers' perspective on the competence children bring to school is situated in the conceptual, praxiological and socio-emotional dimensions of their assessment literacy (Pastore & Andrade, 2019). It is not a straightforward process for teachers to know about the different needs of individual children and adapt their teaching. Recent research on teacher beliefs about inclusion in Norwegian schools has identified how some teachers find inclusion challenging (Aas, 2022). These challenges are connected to teachers' beliefs on teaching and learning that hinder the development of inclusive practice. While assessment in this thesis is connected to both the process and product of learning, teachers' formative use of assessment is always connected to the context of the assessed learning process. A limited view of learning as individualisation as opposed to a social activity, teacher centring in contrast to collaboration between children and teachers, and focus on children's skills without taking the social and emotional context that these skills exist within, cause challenges for teachers' inclusive practice (Aas, 2022). From the perspective of this thesis, we need to consider both the product - what children know, and the process - how we can help children develop their knowledge in a learning community. There is a need for future research to investigate how the situated complexity of teachers' formative assessment affects instructional validity, as teachers' interpretations and use of assessments ultimately affect the facilitation of children's further learning opportunities. One potential avenue is to investigate teachers' use of support materials to identify how these materials can support teachers' assessment literacy development. Another potential avenue is to explore how teachers can use recordings of children's solutions to interactive assessment items in a classroom context. In this way, it might be possible to find connections between the process and product of learning, ensuring that all children are met with sufficient challenges to develop their number competence.

8 References

Aas, H. K. (2022). Teachers talk on student needs: Exploring how teacher beliefs challenge inclusive education in a Norwegian context. *International Journal of Inclusive Education*, *26*(5), 495–509. https://doi.org/10.1080/13603116.2019.1698065

Andrews, P., & Sayers, J. (2015). Identifying opportunities for grade one children to acquire foundational number sense: Developing a framework for cross cultural classroom analyses. *Early Childhood Education Journal*, *43*(4), 257–267. https://doi.org/10.1007/s10643-014-0653-6

Andrews, P., Sayers, J., & Marschall, G. (2015). Developing foundational number sense: Number line examples from Poland and Russia. In K. Krainer & N. Vondrová (Eds.), *Proceedings of the ninth congress of the European Society for Research in Mathematics Education* (pp. 1681–1687). European Society for Research in Mathematics Education. https://hal.science/hal-01287990

Aunio, P., & Räsänen, P. (2016). Core numerical skills for learning mathematics in children aged five to eight years – A working model for educators. *European Early Childhood Education Research Journal*, *24*(5), 684–704. https://doi.org/10.1080/1350293X.2014.996424

Bakker, A., & van Eerde, H. (2015). An introduction to design-based research with an example from statistics education. In A. Bikner-Ahsbahs, C. Knipping, & N. Presmeg (Eds.), *Approaches to qualitative research in mathematics education* (pp. 429–466). Springer. https://doi.org/https://doi.org/10.1007/978-94-017-9181-6

Baroody, A. J., Gannon, K. E., Berent, R., & Ginsburg, H. P. (1984). The development of basic formal mathematics abilities. *Acta Paedologica*, *1*(2), 133–150.

Baumert, J., Kunter, M., Blum, W., Brunner, M., Voss, T., Jordan, A., Klusmann, U., Krauss, S., Neubrand, M., & Tsai, Y.-M. (2010). Teachers' mathematical knowledge, cognitive activation in the classroom, and student progress. *American Educational Research Journal*, *47*(1), 133–180. https://doi.org/10.3102/0002831209345157

Bennett, R. E. (2011). Formative assessment: A critical review. *Assessment in Education: Principles, Policy & Practice*, *18*(1), 5–25. https://doi.org/10.1080/0969594X.2010.513678

Berch, D. B. (2005). Making sense of number sense: Implications for children with mathematical disabilities. *Journal of Learning Disabilities*, *38*(4), 333–339. https://doi.org/10.1177/00222194050380040901

Bergsmo, H. S., Dalvang, T., & Kirsti, L. R. E. (2020). *MIO Matematikken - Individet - Omgivelsene håndbok* (2nd ed.). GAN Aschehoug.

Biesta, G. (2010). Pragmatism and the philosophical foundations of mixed methods research. In A. Tashakkori & C. Teddlie (Eds.), *Sage handbook of mixed methods in social and behavioral research* (Vol. 2, pp. 95–118). Sage. https://doi.org/10.4135/9781506335193

Bisanz, J., Sherman, J. L., Rasmussen, C., & Elaine, H. (2005). Development of arithmetic skills and knowledge in preschool children. In J. I. D. Campbell (Ed.), *The handbook of mathematical cognition* (pp. 143–162). Psychology Press. https://search.ebscohost.com/login.aspx?direct=true&db=nlebk&AN=138710&authtype=sso&custid=ntnu&site=ehost-live

Björklund, C., & Kempe, U. R. (2022). Strategies informed by various ways of experiencing number relations in subtraction tasks. *The Journal of Mathematical Behavior*, *67*, 100994. https://doi.org/10.1016/j.jmathb.2022.100994

Black, P., & Wiliam, D. (2018). Classroom assessment and pedagogy. *Assessment in Education: Principles, Policy & Practice*, *25*(6), 551–575. https://doi.org/10.1080/0969594X.2018.1441807

Blömeke, S., Gustafsson, J.-E., & Shavelson, R. J. (2015). Beyond dichotomies. *Zeitschrift Für Psychologie*, *223*(1), 3–13. https://doi.org/10.1027/2151-2604/a000194

Bodovski, K., & Farkas, G. (2007). Mathematics growth in early elementary school: The roles of beginning knowledge, student engagement, and instruction. In *The Elementary School Journal* (Vol. 108, pp. 115–130). Univ of Chicago Press. https://doi.org/10.1086/525550

Bostic, J. D., Krupa, E. E., Carney, M. B., & Shih, J. C. (2019). Reflecting on the past and looking ahead at opportunities in quantitative measurement of K-12 students' content knowledge. In J. D. Bostic, E. E. Krupa, & J. C. Shih (Eds.), *Quantitative measures of mathematical knowledge: researching instruments and perspectives.* (pp. 205–229). Routledge.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101.

Breadmore, H. L., & Carroll, J. M. (2021). *A systematic review of standardised measures of attainment in literacy, mathematics, and science: Evidence review*.

Brookhart, S. M. (2011). Educational assessment knowledge and skills for teachers. *Educational Measurement: Issues and Practice*, *30*(1), 3–12. https://doi.org/https://doi.org/10.1111/j.1745-3992.2010.00195.x

Brown, A. L. (1992). Design experiments: Theoretical and methodological challenges in creating complex interventions in classroom settings. *The Journal of the Learning Sciences*, *2*(2), 141–178.

Carpenter, T. P., Ansell, E., Franke, M. L., Fennema, E., & Weisbeck, L. (1993). Models of problem solving: A study of kindergarten children's problem-solving processes. *Journal for Research in Mathematics Education JRME*, *24*(5), 428–441. https://doi.org/10.5951/jresematheduc.24.5.0428

Carpenter, T. P., & Moser, J. M. (1984). The acquisition of addition and subtraction concepts in grades one through three. *Journal for Research in Mathematics Education*, *15*(3), 179–202. https://www.jstor.org/stable/748348?seq=1#metadata_info_tab_contents

Clarke, B., Cheeseman, J., & Clarke, D. (2006). The mathematical knowledge and understanding young children bring to school. *Mathematics Education Research Journal*, *18*(1), 78–102. https://doi.org/10.1007/BF03217430

Clarke, D., Clarke, B., & Roche, A. (2011). Building teachers' expertise in understanding, assessing and developing children's mathematical thinking: the power of task-based, one-to-one assessment interviews. *ZDM*, *43*(6), 901–913. https://doi.org/10.1007/s11858-011-0345-2

Clements, D. H., Sarama, J. H., & Liu, X. H. (2008). Development of a measure of early mathematics achievement using the Rasch model: The Research-Based Early Maths

Assessment. *Educational Psychology*, *28*(4), 457–482. https://doi.org/10.1080/01443410701777272

Clements, D. H., Sarama, J., & MacDonald, B. L. (2019). Subitizing: The neglected quantifier. In A. Norton & M. Alibali (Eds.), *Constructing number. Research in Matheamtics Education.* (pp. 13–45). Springer. https://doi.org/10.1007/978-3-030-00491-0_2

Clements, D. H., Sarama, J., Tatsuoka, C., Banse, H., & Tatsuoka, K. (2021). Evaluating a model for developing cognitively diagnostic adaptive assessments: The case of young children's length measurement. *Journal of Research in Childhood Education*, 1–16. https://doi.org/10.1080/02568543.2021.1895921

Clements, D. H., & Sarama, J. (2021). *Learning and teaching early math: The learning trajectories approach* (3rd ed.). Routledge.

Cobb, P., Jackson, K., & Dunlap, C. (2015). Design research: An analysis and critique. In L. D. English & D. Kirshner (Eds.), *Handbook of international research in mathematics education* (pp. 493–515). Routledge.

Collins, A. (1992). Toward a design science of education. In E. Scanlon & T. O'Shea (Eds.), *New directions in educational technology* (pp. 15–22). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-77750-9_2

Collins, M. A., & Laski, E. V. (2015). Preschoolers' strategies for solving visual pattern tasks. *Early Childhood Research Quarterly*, *32*, 204–214. https://doi.org/10.1016/j.ecresq.2015.04.004

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*(4), 281–302. https://doi.org/10.1037/h0040957

Devlin, B. L., Jordan, N. C., & Klein, A. (2022). Predicting mathematics achievement from subdomains of early number competence: Differences by grade and achievement level. *Journal of Experimental Child Psychology*, *217*, 105354. https://doi.org/10.1016/j.jecp.2021.105354

Dowker, A. (2005). *Individual differences in arithmetic implications for psychology, neuroscience and education* (1st ed.). Psychology Press. https://doi.org/10.4324/9780203324899

Dowker, A. (2016). Individual differences in arithmetical abilities: The componential nature of arithmetic. In R. C. Kadosh & A. Dowker (Eds.), *The Oxford handbook of numerical cognition* (pp. 878–894). Oxford University Press.

Drijvers, P. (2018). Digital assessment of mathematics: Opportunities, issues and criteria. *Mesure et Évaluation En Éducation*, *41*(1), 41–66. https://doi.org/10.7202/1055896ar

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., Pagani, L. S., Feinstein, L., Engel, M., Brooks-Gunn, J., Sexton, H., Duckwoth, K., & Japel, C. (2007). School readiness and later achievement. *Developmental Psychology*, *43*(6), 1428–1446. https://doi.org/10.1037/0012-1649.43.6.1428

Ellefsen, I. V. (2021). *Sammenligning av tallforståelseskompetanse hos elever på 1. trinn på data og papir [A comparison of the number sense of first grade students' on paper-based and digital tests]* [[Master thesis] Norwegian University of Science and Technology]. https://hdl.handle.net/11250/2785121

Fischer, C., Pardos, Z. A., Baker, R. S., Williams, J. J., Smyth, P., Yu, R., Slater, S., Baker, R., & Warschauer, M. (2020). Mining big data in education: Affordances and challenges. *Review of Research in Education*, *44*(1), 130–160. https://doi.org/10.3102/0091732X20903304

Gaarden, J. S. (2020). *Strukturell bevissthet i telling og subitizing [Structural awareness in counting and subitizing]*. [Master thesis] Norwegian University of Science and Technology.

Geary, D. C., Bailey, D. H., & Hoard, M. K. (2009). Predicting mathematical achievement and mathematical learning disability with a simple screening tool: The number sets test. *Journal of Psychoeducational Assessment*, *27*(3), 265–279. https://doi.org/10.1177/0734282908330592

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

Ginsburg, H. P. (2016). Helping early childhood educators to understand and assess young children's mathematical minds. *ZDM Mathematics Education*, *48*, 941–946. https://doi.org/10.1007/s11858-016-0807-7

Ginsburg, H. P., & Baroody, A. J. (2003). *TEMA-3: Test of early mathematics ability*. Pro-ed.

Ginsburg, H. P., & Pappas, S. (2016). Invitation to the birthday party: Rationale and description. *ZDM Mathematics Education*, *48*, 947–960. https://doi.org/10.1007/s11858-016-0818-4

Ginsburg, H. P., Wu, R., & Diamond, J. S. (2019). MathemAntics: a model for computer-based mathematics education for young children. *Infancia y Aprendizaje*, *42*(2), 247–302. https://doi.org/10.1080/02103702.2019.1589966

Godino, J. D., Batanero, C., Contreras, A., Estepa, A., Lacasta, E., & Wilhelmi, M. (2013). Didactic engineering as design-based research in mathematics education. In B. Ubuz, C. Haser, & M. A. Mariotti (Eds.), *Proceedings of the Eight Congress of the European Society for Research in Mathematics Education* (pp. 2810–2819). Middle East Technical University and ERME.

Gutiérrez, R. (2012). Context matters: How should we conceptualize equity in mathematics education? In B. Herbel-Eisenmann, J. Choppin, D. Wagner, & D. Pimm (Eds.), *Equity in discourse for mathematics education: Theories, practices, and policies* (pp. 17–33). Springer Netherlands. https://doi.org/10.1007/978-94-007-2813-4_2

Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory : Principles and applications*. Springer Dordrecht. https://doi.org/10.1007/978-94-017-1988-9

Heritage, M., & Heritage, J. (2013). Teacher questioning: The epicenter of instruction and assessment. *Applied Measurement in Education*, *26*. https://doi.org/10.1080/08957347.2013.793190

Heritage, M., & Wylie, C. (2018). Reaping the benefits of assessment for learning: Achievement, identity, and equity. *ZDM*, *50*(4), 729–741. https://doi.org/10.1007/s11858-018-0943-3

Hiebert, J., & Lefevre, P. (1986). Conceptual and procedural knowledge in mathematics: An introductory analysis. In J. Hiebert (Ed.), *Conceptual and Procedural Knowledge : The Case of Mathematics* (pp. 1–23). Taylor & Francis Group. http://ebookcentral.proquest.com/lib/ntnu/detail.action?docID=1356003

Hodgen, J., Geraniou, E., Bolondi, G., & Ferretti, F. (2022). *Proceedings of the twelfth congress of the European society for research in mathematics education (CERME12)*. https://hal.science/CERME12/search/index/?q=%2A&domain_t=math

Hodgen, J. (2011). Knowing and identity: A situated theory of mathematics knowledge in teaching. In T. Rowland & K. Ruthven (Eds.), *Mathematical knowledge in teaching* (pp. 27–42). Springer Netherlands. https://doi.org/10.1007/978-90-481-9766-8_3

Holgersson, I., Barendregt, W., Emanuelsson, J., Ottosson, T., Rietz, E., & Lindström, B. (2016). Fingu—A game to support children's development of arithmetic competence: Theory, design and empirical research. In *International perspectives on teaching and learning mathematics with virtual manipulatives* (pp. 123–145). Springer.

Holvik, S. B. (2022). *Utvikling av elevar si grunnleggande talforståing [Development of students' foundational number sense]*. [Master thesis] Norwegian University of Science and Technology.

Imsen, G., & Volckmar, N. (2014). The Norwegian school for all: Historical emergence and neoliberal confrontation. In U. Blossing, G. Imsen, & L. Moos (Eds.), *The Nordic education model: "A School for all" Encounters Neo-Liberal Policy* (pp. 35–55). Springer Netherlands. https://doi.org/10.1007/978-94-007-7125-3_3

Inkscape Project. (2023). *Inkscape*. https://inkscape.org/

Johnson, B. R., Onwuegbuzie, A. J., & Turner, L. A. (2007). Toward a definition of mixed methods research. *Journal of Mixed Methods Research*, *1*(2), 112–133. https://doi.org/10.1177/1558689806298224

Jordan, N. C., Kaplan, D., Ramineni, C., & Locuniak, M. N. (2009). Early math matters: Kindergarten number competence and later mathematics outcomes. *Developmental Psychology*, *45*(3), 850–867. https://doi.org/10.1037/a0014939

Jordan, N. C., Devlin, B. L., & Botello, M. (2022). Core foundations of early mathematics: Refining the number sense framework. *Current Opinion in Behavioral Sciences*, *46*, 101181. https://doi.org/10.1016/j.cobeha.2022.101181

Jordan, N. C., Glutting, J., Ramineni, C., & Watkins, M. W. (2010). Validating a number sense screening tool for use in kindergarten and first grade: Prediction of mathematics proficiency in third grade. *School Psychology Review*, *39*(2), 181–195. https://doi.org/10.1080/02796015.2010.12087772

Kane, M. (2016). Explicating validity. *Assessment in Education: Principles, Policy & Practice*, *23*(2), 198–211. https://doi.org/10.1080/0969594X.2015.1060192

Kane, M., Crooks, T., & Cohen, A. (1999). Validating measures of performance. *Educational Measurement: Issues and Practice*, *18*(2), 5–17. https://doi.org/10.1111/j.1745-3992.1999.tb00010.x

Kilpatrick, J. (2014). Competency frameworks in mathematics education. In S. Lerman (Ed.), *Encyclopedia of mathematics education* (pp. 85–87). Springer Netherlands. https://doi.org/10.1007/978-94-007-4978-8_27

Kilpatrick, J., Swafford, J., & Findell, B. (2001). *Adding it up: Helping children learn mathematics.* National Academies Press.

Koeppen, K., Hartig, J., Klieme, E., & Leutner, D. (2008). Current issues in competence modeling and assessment. *Zeitschrift Für Psychologie [Journal of Psychology]*, *216*(2), 61–73. https://doi.org/10.1027/0044-3409.216.2.61

Krauss, S., Brunner, M., Kunter, M., Baumert, J., Blum, W., Neubrand, M., & Jordan, A. (2008). Pedagogical content knowledge and content knowledge of secondary mathematics teachers. *Journal of Educational Psychology, 100*(3), 716–725. https://doi.org/10.1037/0022-0663.100.3.716

Krupa, E. E., Bostic, J. D., & Shih, J. C. (2019). Validation in mathematics education: An introduction to quantitative measures of mathematical knowledge: Researching instruments and perspectives. In E. E. Krupa, J. D. Bostic, & J. C. Shih (Eds.), *Quantitative measures of mathematical knowledge researching instruments and perspectives* (pp. 1–13). Routledge.

Le Corre, M., & Carey, S. (2007). One, two, three, four, nothing more: An investigation of the conceptual sources of the verbal counting principles. *Cognition*, *105*(2), 395–438. https://doi.org/10.1016/J.COGNITION.2006.10.005

LeFevre, J. A., Smith-Chant, B. L., Fast, L., Skwarchuk, S. L., Sargla, E., Arnup, J. S., Penner-Wilger, M., Bisanz, J., & Kamawar, D. (2006). What counts as knowing? The development of conceptual and procedural knowledge of counting from kindergarten through grade 2. *Journal of Experimental Child Psychology*, *93*(4), 285–303. https://doi.org/10.1016/j.jecp.2005.11.002

Lembke, E., & Foegen, A. (2009). Identifying early numeracy indicators for kindergarten and first-grade students. *Learning Disabilities Research & Practice*, *24*(1), 12–20.

Linacre, J. M. (2017). *Winsteps® Rasch measurement computer program.* Beaverton, OR: Winsteps.com.

Linacre, J. M. (2022). *A user's guide to Winsteps® Ministep Rasch model computer programs Program manual 5.3.4*. https://www.winsteps.com/manuals.htm

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Sage.

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

Lopez-Pedersen, A., Mononen, R., Korhonen, J., Aunio, P., & Melby-Lervåg, M. (2021). Validation of an early numeracy screener for first graders. *Scandinavian Journal of Educational Research*, *65*(3), 404–424. https://doi.org/10.1080/00313831.2019.1705901

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

McIntosh, A. (2007). *Alle teller! Håndbok for lærere som underviser i matematikk i grunnskolen [Everyone counts! Handbook for mathematics teachers in primary school]* (I. Stedøy & M. R. Settemsdal (eds.)). Matematikksenteret.

McIntosh, A., Reys, B. J., & Reys, R. E. (1992). A proposed framework for examining basic number sense. *For the Learning of Mathematics*, *12*(3), 2–44. https://www.jstor.org/stable/pdf/40248053.pdf

Messick, S. (1989). Meaning and values in test validation: The science and ethics of assessment. *Educational Researcher*, *18*(2), 5–11. https://doi.org/https://doi.org/10.2307/1175249

Middletion, J., Gorard, S., Taylor, C., & Bannan-Ritland, B. (2008). The "Compleat" Design Experiment From Soup to Nuts. In A. E. Kelly, R. A. Lesh, & J. Y. Baek (Eds.), *Handbook of design research methods in education: Innovations in science, technology, engineering, and mathematics learning and teaching* (pp. 21–46). Routledge.

Milburn, T. F., Lonigan, C. J., DeFlorio, L., & Klein, A. (2019). Dimensionality of preschoolers' informal mathematical abilities. *Early Childhood Research Quarterly*, *47*, 487–495. https://doi.org/https://doi.org/10.1016/j.ecresq.2018.07.006

Morgan, D. L. (2007). Paradigms lost and pragmatism regained: Methodological implications of combining qualitative and quantitative methods. *Journal of Mixed Methods Research*, *1*(1), 48–76.

Moss, P. A. (2013). Validity in action: Lessons from studies of data use. *Journal of Educational Measurement*, *50*(1), 91–98. https://doi.org/10.1111/jedm.12003

Moss, P. A., Girard, B. J., & Haniford, L. C. (2006). Validity in educational assessment. *Review of Research in Education*, *30*(1), 109–162. https://doi.org/10.3102/0091732X030001109

National agency for education. (2019). *Hitta matematiken: Nationellt kartläggingsmaterial i matematiskt tänkande i förskolaklass [Find the mathematics: National assessment material in mathematical thinking in preschool class].*

Neuwirt, C. (2021). *Designing a digital math-placement-test for Norwegian preschool children*. FH Johanneum Graz.

Niss, M., Bruder, R., Planas, N., Turner, R., & Villa-Ochoa, J. A. (2016). Survey team on: Conceptualisation of the role of competencies, knowing and knowledge in mathematics education research. *ZDM Mathematics Education*, *48*(5), 611–632. https://doi.org/10.1007/s11858-016-0799-3

Nortvedt, G. A., Buchholtz, N., Fauskanger, J., Hähkiöniemi, M., Jessen, B. E., Naalsund, M., Nilsen, H. K., Pálsdóttir, G., Portaankorva-Koivisto, P., Radišić, J. Ö., Sigurjónsson, J., Viirman, O., & Wernberg, A. (2022). *Bringing Nordic mathematics education into the future. Proceedings of Norma 20: The ninth Nordic conference on mathematics education*. http://matematikdidaktik.org/wpcontent/%0Auploads/1672/22/17_NORMA_20_proceedings.pdf

Nortvedt, G. A, & Buchholtz, N. (2018). Assessment in mathematics education: responding to issues regarding methodology, policy, and equity. *ZDM*, *50*. https://doi.org/10.1007/s11858-018-0963-z

Norwegian directorate for education and training. (2020). *Læreplan i Matematikk 1–10 (MAT01-05), kompetansemål og vurdering*. https://www.udir.no/lk20/mat01-05/kompetansemaal-og-vurdering/kv20

Norwegian directorate for education and training. (2023). *Kartleggingsprøver i regning*. https://www.udir.no/eksamen-og-prover/prover/kartlegging-gs/

Olsher, S., Yerushalmy, M., & Chazan, D. (2016). How might the use of technology in formative assessment support changes in mathematics teaching? *For the Learning of Mathematics*, *36*(3), 11–18.

Pastore, S., & Andrade, H. L. (2019). Teacher assessment literacy: A three-dimensional model. *Teaching and Teacher Education*, *84*, 128–138. https://doi.org/10.1016/j.tate.2019.05.003

Pellegrino, J. W., Chudowsky, N., & Glaser, R. (2001). *Knowing what students know : the science and design of educational assessment*. National Academy Press.

Pellegrino, J. W., DiBello, L. V, & Goldman, S. R. (2016). A framework for conceptualizing and evaluating the validity of instructionally relevant assessments. *Educational Psychologist*, *51*(1), 59–81. https://doi.org/10.1080/00461520.2016.1145550

Phillips, D., & Crowell, N. A. (1994). *Cultural diversity and early education: Report of a workshop*. https://files.eric.ed.gov/fulltext/ED376982.pdf

Piaget, J. (1970). *Genetic epistemology*. Columbia University Press.

Popham, W. J. (2009). Assessment literacy for teachers: Faddish or fundamental? *Theory Into Practice*, *48*(1), 4–11. https://doi.org/10.1080/00405840802577536

Proulx, J. (2013). Mental mathematics, emergence of strategies, and the enactivist theory of cognition. *Educational Studies in Mathematics*, *84*(3), 309–328. https://doi.org/10.1007/s10649-013-9480-8

Purpura, D. J., & Lonigan, C. J. (2013). Informal Numeracy Skills: The Structure and Relations Among Numbering, Relations, and Arithmetic Operations in Preschool. *American Educational Research Journal*, *50*(1), 178–209. https://doi.org/10.3102/0002831212465332

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. University of Chicago Press.

Raudenbush, S. W., Hernandez, M., Goldin-Meadow, S., Carrazza, C., Foley, A., Leslie, D., Sorkin, J. E., & Levine, S. C. (2020). Longitudinally adaptive assessment and instruction increase numerical skills of preschool children. *Proceedings of the National Academy of Sciences*, *117*(45), 27945–27953. https://doi.org/10.1073/pnas.2002883117

Rittle-Johnson, B., & Star, J. R. (2007). Does comparing solution methods facilitate conceptual and procedural knowledge? An experimental study on learning to solve equations. *Journal of Educational Psychology*, *99*(3), 561–574. https://doi.org/10.1037/0022-0663.99.3.561

Rost, J. (2001). The growing family of Rasch models. In A. Boomsma, M. A. J. van Duijn, & T. A. B. Snijders (Eds.), *Essays on item response theory* (pp. 25–42). Springer New York. https://doi.org/10.1007/978-1-4613-0169-1_2

Saksvik-Raanes, G., & Solstad, T. (2022). Factors that influence the difficulty level of digital arithmetic assessment items for first-grade students. *Proceedings of the 12th Congress of the European Society for Research in Mathematics Education (CERME12), Feb 2022*. hal-03753463

Saksvik-Raanes, G., & Solstad, T. (2023). *Validity evidence for the formative assessment of early number sense with digital tools and interactive items [submitted to ESM]*.

Saksvik-Raanes, G., & Solstad, T. (2021). Developing a formative, teacher-oriented, digital tool to assess number sense in school starters. In G. A Nortvedt, N. F. Buchholtz, J. Fauskanger, M. Hähkiöniemi, B. E. Jessen, M. Naalsund, H. K. Nilsen, G. Pálsdóttir, P. Portaankorva-Koivisto, J. Radišić, J. Ö. Sigurjónsson, O. Viirman, & A. Wernberg (Eds.), *Bringing Nordic mathematics education into the future Proceedings of Norma 20 The ninth Nordic Conference on Mathematics Education* (pp. 212–219). Swedish Society for Research in Mathematics Education.

Saksvik-Raanes, G., Solstad, T., & Grimeland, Y. (2023). Et innblikk i barns tallforståelse på slutten av barnehageårene gjennom digitale oppgaver [A survey of five-year-old children’s number sense in early childhood education using digital]. *Nordisk Barnehageforskning*, *20*(2), 5–28. https://doi.org/10.23865/nbf.v20.279

Sayers, J., Andrews, P., & Boistrup, L. B. (2016). The role of conceptual subitising in the development of foundational number sense. In W. A. Meaney T., Helenius O., Johansson M., Lange T. (Ed.), *Mathematics education in the early years* (pp. 371–394). Springer. https://doi.org/10.1007/978-3-319-23935-4_21

Schjølberg, A. (2021). *“Jeg må prøve å telle” En kvalitativ studie om elevers strategier i arbeid med tallforståelsesoppgaver på første trinn ["I need to try to count" A qualitative study on students’ strategies working with number sense tasks in first grade]*. [Master thesis] Norwegian University of Science and Technology.

Schumacker, R. E. (2004). Rasch measurement: The dichotomus model. In J. E. V Smith & M. Smith, R (Eds.), *Introduction to Rasch measurement theory, models and applications* (pp. 226–257). JAM Press.

Siegler, R. S., & Booth, J. L. (2004). Development of numerical estimation in young children. *Child Development*, *75*(2), 428–444.

Sinclair, N., Chorney, S., & Rodney, S. (2016). Rhythm in number: Exploring the affective, social and mathematical dimensions of using TouchCounts. *Mathematics Education Research Journal*, *28*(1), 31–51. https://doi.org/10.1007/s13394-015-0154-y

Sinclair, N., & Jackiw, N. (2014). *TouchCounts [software application for the iPad]*. https://itunes.apple.com/ca/app/touchcounts/id897302197?mt=8

Sjoe, N. M., Bleses, D., Dybdal, L., Tideman, E., Kirkeby, H., Sehested, K. K., Nielsen, H., Kreiner, S., & Jensen, P. (2019). Short Danish version of the tools for early assessment in math (TEAM) for 3–6-year-olds. *Early Education and Development*, *30*(2), 238–258. https://doi.org/10.1080/10409289.2018.1544810

Skemp, R. R. (1976). Relational understanding and instrumental understanding. *Mathematics Teaching*, *77*(1), 20–26.

Smith, R. M., & Wind, S. (2018). *Rasch measurement models: Interpreting WINSTEPS and FACETS output* (2nd ed.). JAM Press.

Smith, J. E. V. (2004). Detecting and evaluating the impact of multidimensionality using item fit statistics and principal component analysis of residuals. In J. E. V Smith & R. M. Smith (Eds.), *Introduction to Rasch Measurement Theory, models and applications* (pp. 575–599). JAM Press.

Søberg, V. (2020). *“Jeg bare telte, jeg” En kvalitativ forskningsstudie om elevers bruk av tellestrategier i digitale oppgavesett. [I just counted! A qualitative study on students’* *counting strategies in digital tasks]*. [Master thesis] Norwegian University of Science and Technology.

Soto, M., & Ambrose, R. (2016). Screencasts: Formative assessment for mathematical thinking. *Technology, Knowledge and Learning*, *21*(2), 277–283. https://doi.org/10.1007/s10758-015-9272-6

Sowder, J. T., & Schappelle, B. P. (1989). *Establishing foundations for research on number sense and related topics: Report of a conference*.

Star, J. R. (2005). Reconceptualizing procedural knowledge. *Journal for Research in Mathematics Education*, *36*(5), 404–411. https://doi.org/10.2307/30034943

ten Braak, D., & Størksen, I. (2021). Psychometric properties of the Ani Banani math yest. *European Journal of Developmental Psychology*, *18*(4), 610–628. https://doi.org/10.1080/17405629.2021.1879046

Threlfall, J. (2002). Flexible mental calculation. *Educational Studies in Mathematics*, *50*(1), 29–47. https://doi.org/10.1023/A:1020572803437

Thurstone, L. L. (1959). *The measurement of values*. University of Chicago Press.

Tucker, S. I., & Johnson, T. N. (2022). Developing number sense with Fingu: A preschooler's embodied mathematics during interactions with a multi-touch digital game. *Mathematics Education Research Journal*, *34*(2), 393–417. https://doi.org/10.1007/s13394-020-00349-4

Urban, M., Reikerås, E., Eidsvåg, Gunnar M. Guevara, J., Saegø, J., & Semmoloni, C. (2022). *Nordic approaches to evaluation and assessment in early childhood education and care*. http://dx.doi.org/10.6027/temanord2022-512

Van de Rijt, B. A. M., Van Luit, J. E. H., & Pennings, A. H. (1999). The construction of the Utrecht early mathematical competence scales. *Educational and Psychological Measurement*, *59*(2), 289–309.

van de Rijt, B., van Luit, J. E. H., & Pennings, A. H. (1999). The Construction of the Utrecht early mathematical competence scales. *Educational and Psychological Measurement*, *59*(2), 289–309. https://doi.org/10.1177/0013164499592006

van den Heuvel-Panhuizen, M., Kolovou, A., & Peltenburg, M. (2011). Using ICT to Improve Assessment. In B. Kaur & K. Y. Wong (Eds.), *Assessment In The Mathematics Classroom: Yearbook 2011, Association of Mathematics Educators* (pp. 165–185). World Scientific. https://doi.org/https://doi.org/10.1142/9789814360999_0007

van Den Heuvel-Panhuizen, M. (2005). The role of contexts in assessment problems in mathematics. *For the Learning of Mathematics*, *25*(2), 2–23.

Verschaffel, L., Luwel, K., Torbeyns, J., & Van Dooren, W. (2009). Conceptualizing, investigating, and enhancing adaptive expertise in elementary mathematics education. *European Journal of Psychology of Education*, *24*(3), 335–359. https://doi.org/10.1007/BF03174765

Vygotskij, L. S. (1978). *Mind in society : The development of higher psychological processes*. Harvard University Press.

Wager, A. A., & Parks, A. N. (2016). Assessing early number learning in play. *ZDM*, *48*(7), 991–1002. https://doi.org/10.1007/s11858-016-0806-8

Walker, R. (2011). Design-based research: Reflections on some epistemological issues and practices. In L. Markauskaite, P. Freebody, & J. Irwin (Eds.), *Methodological choice and design* (pp. 51–56). Springer. https://doi.org/10.1007/978-90-481-8933-5

Weaver, K. (2018). Pragmatic paradigm. In B. B. Frey (Ed.), *The SAGE encyclopedia of educational research, measurement, and evaluation* (pp. 1287–1288). Sage Publications. https://doi.org/10.4135/9781506326139

Whitacre, I., Henning, B., & Atabaș, Şebnem. (2020). Disentangling the research literature on number sense: Three constructs, one name. *Review of Educational Research*, *90*(1), 95–134. https://doi.org/10.3102/0034654319899706

Wiliam, D. (2018). Assessment for learning: Meeting the challenge of implementation. *Assessment in Education: Principles, Policy & Practice*, *25*(6), 682–685. https://doi.org/10.1080/0969594X.2017.1401526

Wilkins, J. L. M., MacDonald, B. L., & Norton, A. (2022). Construction of subitized units is related to the construction of arithmetic units. *Educational Studies in Mathematics*, *109*(1), 137–154. https://doi.org/10.1007/s10649-021-10076-7

Wilson, M. (2005). *Constructing measures: An item response modeling approach*. Routledge. https://doi.org/10.4324/9781410611697

Wolfe, E. W., & Smith, E. V. (2007a). Instrument development tools and activities for measure validation using Rasch models: Part I instrument development tools. *Journal of Applied Measurement*.

Wolfe, E. W., & Smith, J. E. V. (2007b). Instrument development tools and activities for measure validation using Rasch models: Part II validation activities. *Journal of Applied Measurement*, *8*(2), 204–234.

Wright, B. D. (1977). Solving measurement problems with the Rasch model. *Journal of Educational Measurement*, *14*(2), 97–116. http://www.jstor.org/stable/1434010

Wright, B. D., & Mok, M. M. C. (2004). An overview of the family of Rasch measurement models. In E. V. Smith & R. M. Smith (Eds.), *Introduction to Rasch measurement Theory, models and applications* (pp. 1–24). JAM Press.

Wright, B., & Linacre, J. (1994). Reasonable mean-square fit values. *Rasch Measurement Transactions*, *8*(3), 370.

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

Xu, Y., & Brown, G. T. L. (2016). Teacher assessment literacy in practice: A reconceptualization. *Teaching and Teacher Education*, *58*, 149–162. https://doi.org/10.1016/j.tate.2016.05.010

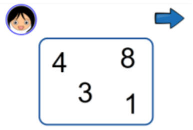
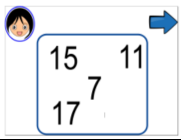
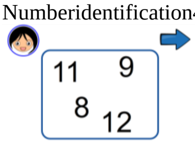
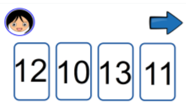
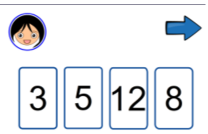
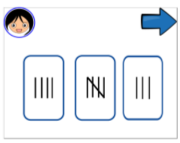
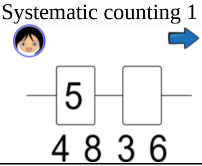
Yilmaz, K. (2013). Comparison of quantitative and qualitative research traditions: Epistemological, theoretical, and methodological differences. *European Journal of Education*, *48*(2), 311–325. https://doi.org/10.1111/ejed.12014

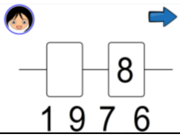
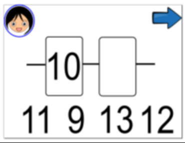
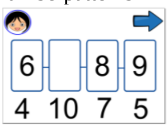
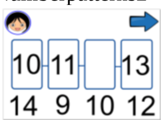
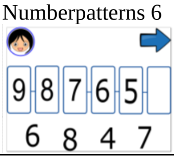
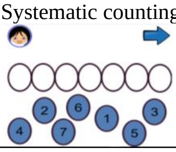
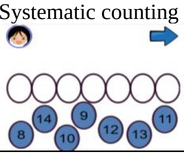
Zhu, Y. (2018). Equity in mathematics education: What did TIMSS and PISA Tell us in the last two decades? In G. Kaiser, H. Forgasz, M. Graven, A. Kuzniak, E. Simmt, & B. Xu

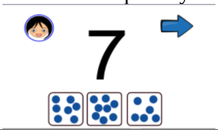
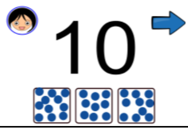
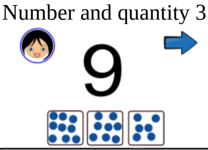
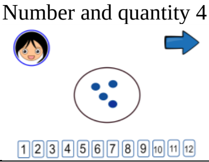
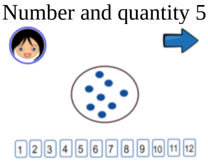
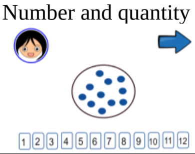
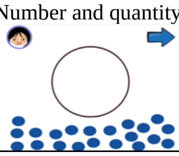
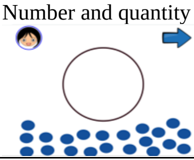
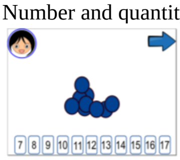
(Eds.), *Invited lectures from the 13th international congress on mathematical education* (pp. 769–786). Springer International Publishing.

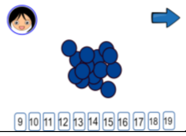
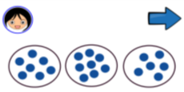
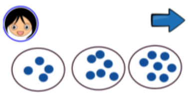
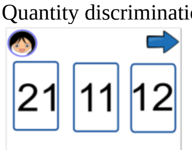
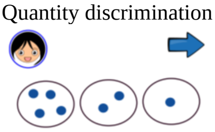
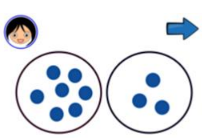
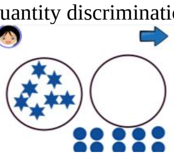
9 Appendicies

Appendix A: Overview of ENSA

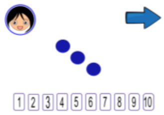
Item	Content
Numberidentification1 	"Press the number…." 4-**3**-8-1
Numberidentification3 	15-7-11-**17**
Numberidentification4 	**11**-13-14-12
Numberidentification5 	12-10-**13**-11
Numberidentification6 	**6**-3-7-9
Numberidentification7 	3-5-12-**8**
Numberidentification9 	**4**-5-3
Systematic counting 1 	"Finish the pattern." 5-**6**
Systematic counting 2	**7**-8

	
Systematic counting 3 	10-**11**
Numberpatterns1 	6-**7**-8-9
Numberpatterns2 	10-11-**12**-13
Numberpatterns 6 	9-8-7-6-**4**
Systematic counting 4-1 	"Order the blue circles with the numerals by drag-and-drop onto the correct target (white circles)." 1-7
Systematic counting 4-2 	8-14
Systematic counting 5 	"Press the … star in the row." 2.
Systematic counting 6 	5.
Systematic counting 7	"Move the blue star on top of the …star in the row." 3.

	
Number and quantity 1 	“Press the box where there are the same as the number shows.” 7
Number and quantity 2 	10
Number and quantity 3 	9
Number and quantity 4 	“Press the numeral that shows how many there are inside the circle.” 6
Number and quantity 5 	8
Number and quantity 6 	11
Number and quantity 7 	“Drag and drop (…) blue objects into the large circle.” 7
Number and quantity 8 	10
Number and quantity 11 	“Move the blue objects to count them and press the numeral corresponding to the number of objects counted.” 9

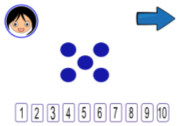
Number and quantity 10  	16
Quantity discrimination 1 	“Press where there are most.”
Quantity discrimination 2 	“…. fewest.”
Quantity discrimination 3 	“…more than 4.”
Quantity discrimination 4  	“…least.” 9-**6**-8
Quantity discrimination 4-1 (estimation 4)  	“…most.” **21**-11-12
Quantity discrimination 5 	“…more than 3.” 4-**2**-1
Quantity discrimination 6 	“Drag and drop the blue objects so that there is an equal amount in each circle.”
Quantity discrimination 7 	

Subitising 1



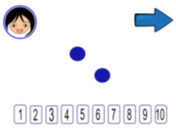
"Press the numeral that corresponds to the number of objects."
3

Subitising 2



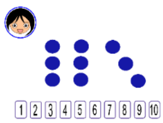
5

Subitising 3



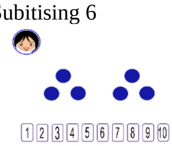
2

Subitising 5



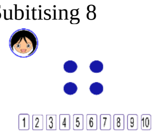
6 + 3

Subitising 6



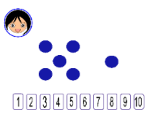
3 + 3

Subitising 8



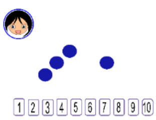
4

Subitising 9



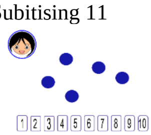
5 + 1

Subitising 10



3 + 1

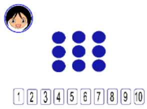
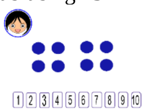
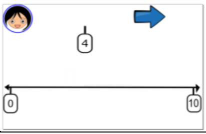
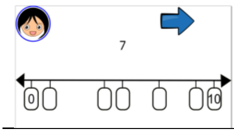
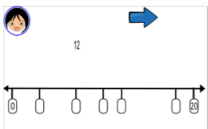
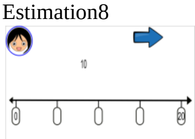
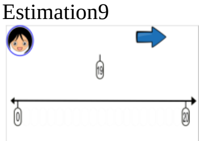
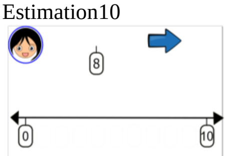
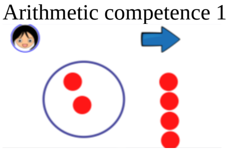
Subitising 11

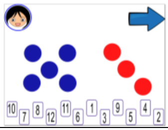


5

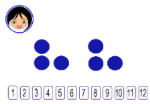
Subitising 12

3+3+3

	
Subitising 13 	4 + 4
Estimation5 	"Move the numeral to the right position on the line."
Estimation6 	
Estimation7 	
Estimation8 	
Estimation9 	
Estimation10 	
Arithmetic competence 1 	“Drag and drop the red objects to end up with four objects inside the blue circle.” 2+_=4
Arithmetic competence 3	“Press the number that shows how many there are all together.” 5+3=8

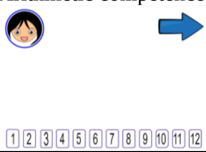


Arithmetic competence 4



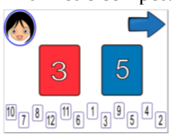
"You have three marbles and get three more; how many do you have now?"
3+3=6

Arithmetic competence 5



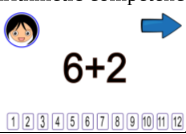
"You have two marbles and get five more; how many do you have now?"
2+5=7

Arithmetic competence 7



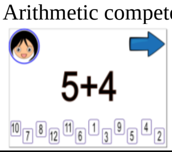
"There are three marbles in the red box. In the blue, there are five. They should have the same. How many more go into the red box?"
5-3=2

Arithmetic competence 8



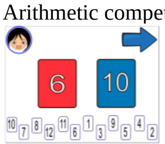
6+2=8

Arithmetic competence 9



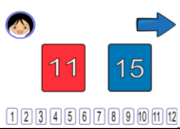
5+4=9

Arithmetic competence 10



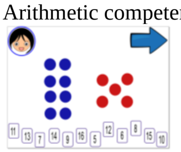
6+x=10
"There are 6 marbles in the red box. There are 10 marbles in the blue box. How many more are there in the blue box?"

Arithmetic competence 11



11+x=15
"There are 11 marbles in the red box. In the blue, there are 15. They should have the same. How many more go into the red box?"

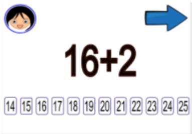
Arithmetic competence 12



8+5=13

Arithmetic competence 13

16+2=18

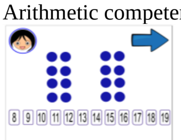


Arithmetic competence 14

11+4=15

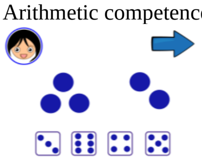
Arithmetic competence 15

8+8=16



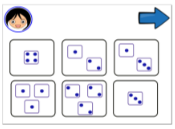
Arithmetic competence 2

“Press the dice that show how many you see.”



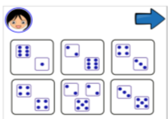
Representingnumber1

«Press where the dice that makes three.»
1+2=1+1+1
3



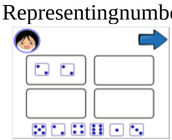
Representingnumber2

«Press where the dice that makes eight.»
2+6=4+4
3+5



Representingnumber4

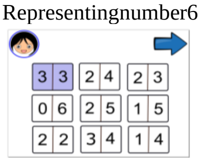
“Drag and drop dice to compose a different set of …dots in each frame.”
2+2=3+1
4=1+1+1+1



Representingnumber6

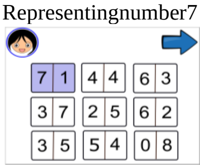
“Find numerals that equal six.”

3+3=2+4
0+6=1+5



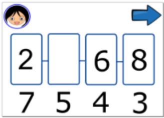
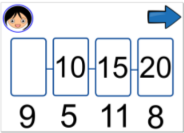
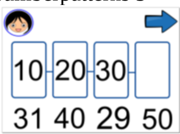
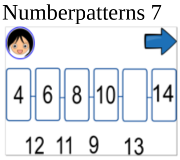
Representingnumber7

“…eight”
7+1=4+4
6+2=3+5
0+8



Numberpatterns3

"Finish the pattern."
2-**4**-6-8

	
Numberpatterns 4 	**5**-10-15-20
Numberpatterns 5 	10-20-30-**40**
Numberpatterns 7 	4-6-8-10-**12**-14

Appendix B: Letters of consent

NTNU

Deltakelse i forskningsprosjektet *Digitale verktøy og tallforståelse*

Dette er et spørsmål til dere foreldre/foresatte om deres barn kan delta i et forskningsprosjekt om tallforståelse på 1. trinn. I dette skrivet gir vi dere informasjon om målene for prosjektet og hva deltakelse vil innebære for deres barn. På siste side er det en samtykkeerklæring som må returneres til skolen hvis dere gir tillatelse til behandling av personidentifiserende informasjon vedrørende deres barn i prosjektet.

Formål

Formålet med prosjektet er å lage et digitalt verktøy som skal hjelpe lærerne å vurdere elevers tallforståelse på 1. trinn. Verktøyet består av en samling tallforståelsesoppgaver som barna skal jobbe med på nettbrett. Oppgavene skal være motiverende og tilpasset elevenes forutsetninger. Hensikten er at elevene, gjennom å gjøre disse oppgavene, skal få vise hva de kan om tall slik at lærerne kan ha mulighet til å tilpasse matematikkundervisningen i det videre arbeidet på skolen.

Hvem er ansvarlig for forskningsprosjektet?

NTNU, Institutt for lærerutdanning ved doktorgradsstipendiat Gunnhild Saksvik-Raanes.

Hvorfor får du spørsmål om å delta?

NTNU har avtale med rektor og lærere på deres 1. trinn om deltakelse i dette prosjektet. Det er oppgavene barna jobber med som skal vurderes i studien. Dermed behøver vi ikke samle inn noen personidentifiserende data vedrørende barnet. Samtidig kan vi ha et behov for å følge elevenes utvikling utover skoleåret, for å vurdere om verktøyet kan si noe om elevenes læringsprosess. For å gjøre dette må vi ta vare på elevens navn for å knytte dette til navnet fra en senere besvarelse på de samme tallforståelsesoppgavene. Det vil også være nyttig for utviklingen av verktøyet å ta opp noe av arbeidet med oppgavene på video.

Hva innebærer det for deg å delta?

For elevene vil deltakelse i prosjektet være del av en vanlig skoledag. De vil gjøre matematikkoppgaver på nettbrett på et grupperom med flere av sine medelever, i tillegg til den ansvarlige for prosjektet. Alle elever skal oppleve mestring under arbeidet og vil ha mulighet til å få hjelp hvis de har behov for det. Noe av arbeidet med oppgavene kan tas opp på video. Dette gjelder kun for de elever der foresatte har samtykket til dette. Alle elever på trinnet vil ha mulighet til å delta i prosjektet uten at det samles inn noen personidentifiserende data om dem.

Det er frivillig å delta
Det er frivillig å delta i prosjektet. Hvis dere velger at barnet skal delta, kan dere når som helst trekke samtykke tilbake uten å oppgi noen grunn. Alle opplysninger om eleven vil da bli anonymisert. Det vil ikke ha noen negative konsekvenser for dere hvis dere ikke vil delta eller senere velger å trekke samtykket.

Personvern – hvordan vi oppbevarer og bruker opplysningene

Vi vil bare bruke opplysningene til formålene vi har fortalt om i dette skrivet. Vi behandler opplysningene konfidensielt og i samsvar med personvernregelverket.

De som vil ha tilgang til materialet er prosjektleder, masterchildrener, forskningsgruppe og veiledere Trygve Solstad og Yvonne Grimeland.

Videomaterialet vil under prosjektet oppbevares på et sikkert sted og i transkripsjonene fra videofilmene vil elevene anonymiseres.

Elevene som deltar, vil ikke kunne gjenkjennes i publikasjoner av prosjektet.

Hva skjer med opplysningene når vi avslutter forskningsprosjektet?

Prosjektet skal etter planen avsluttes i mai 2023. Datamaterialet fra prosjektet vil da anonymiseres. Det anonymiserte materialet vil da kunne oppbevares for videre forskningsarbeid.

Dine rettigheter

Så lenge barnet kan identifiseres i datamaterialet, har dere rett til:

innsyn i hvilke personopplysninger som er registrert,

å få rettet personopplysninger om barnet,

få slettet personopplysninger om barnet,

få utlevert en kopi av barnets personopplysninger og

å sende klage til personvernombudet eller Datatilsynet om behandlingen av personopplysninger.

Covid-19

Under arbeidet med oppgavene forholder vi oss til gjeldende retningslinjer med tanke på smittevern. Vi desinfiserer nettbrett, bord og berøringspunkter, sørger for avstand og håndvask.

Hva gir oss rett til å behandle personopplysninger om ditt barn?

Vi behandler opplysninger om ditt barn basert på ditt samtykke.

På oppdrag fra NTNU har NSD – Norsk senter for forskningsdata AS vurdert at behandlingen av personopplysninger i dette prosjektet er i samsvar med personvernregelverket.

Hvor kan jeg finne ut mer?

Hvis du har spørsmål til studien, eller ønsker å benytte deg av dine rettigheter, ta kontakt med:

Gunnhild Saksvik-Raanes

Doktorgradsstipendiat

Institutt for lærerutdanning

NTNU

7491, Trondheim, Norway

Tlf: 41621593

Veileder for prosjektet: Trygve Solstad

Vårt personvernombud: Thomas Helgesen ved NTNU

NSD – Norsk senter for forskningsdata AS, på epost personverntjenester@nsd.no eller telefon: 55 58 21 17.

Med vennlig hilsen

Gunnhild Saksvik-Raanes

Prosjektansvarlig

NTNU

Deltakelse i forskningsprosjekt om *Digitale verktøy og tallforståelse*

Dette er et spørsmål til deg som lærer om å delta i et forskningsprosjekt hvor formålet er å forbedre elevers første møte med tall i skolesammenheng. I dette skrivet gir vi deg informasjon om målene for prosjektet og hva deltakelse vil innebære for deg.

Formål

Formålet med prosjektet er å lage et digitalt verktøy for å gjøre det lettere for lærere å vurdere elevers tallforståelse. I utviklingen av dette verktøyet vil elever på 1. trinn få prøve ulike matematikkoppgaver på nettbrett i løpet av undervisningen i skolen. Vi ønsker også å finne ut mer om lærernes perspektiver på dette verktøyet og generelt om arbeidet med tallforståelse på 1. trinn.

Hvem er ansvarlig for forskningsprosjektet?

NTNU Institutt for lærerutdanning ved doktorgradsstipendiat Gunnhild Saksvik-Raanes.

Hvorfor får du spørsmål om å delta?

NTNU og ledelsen ved deres skole har avtale om samarbeid i dette prosjektet.

Hva innebærer det for deg å delta?

Deltakelse i prosjektet for deg vil innebære å delta i en uformell samtale (tidsrom ca. 45 min) med prosjektansvarlig der du får presentert resultater fra en tallforståelseskartlegging gjennomført av to av dine elever. I samtalen med prosjektansvarlig vil du få anledning til å komme med dine refleksjoner rundt elevens resultater og på hvilken måte du tenker dette gir kunnskap du trenger for å tilpasse videre undervisning. Vi vil på forhånd ha fått samtykke fra elevenes foresatte til å gjennomføre kartleggingen og samtalen. Samtalen vil bli tatt opp på lydfil og senere transkribert og anonymisert.

Det er frivillig å delta
Det er frivillig å delta i prosjektet og du kan når som helst trekke samtykke tilbake uten å oppgi noen grunn. Alle opplysninger om deg vil da bli anonymisert. Det vil ikke ha noen negative konsekvenser for deg hvis du ikke vil delta eller senere velger å trekke samtykket.

Personvern – hvordan vi oppbevarer og bruker opplysningene

Vi vil bare bruke opplysningene til formålene vi har fortalt om i dette skrivet. Vi behandler opplysningene konfidensielt og i samsvar med personvernregelverket.

De som vil ha tilgang til materialet er prosjektleder, masterchildrener, forskningsgruppe og veiledere Trygve Solstad og Yvonne Grimeland.

Deltakere vil ikke kunne gjenkjennes i publikasjoner av prosjektet.

Hva skjer med opplysningene når vi avslutter forskningsprosjektet?

Prosjektet skal etter planen avsluttes i mai 2023. Datamaterialet fra prosjektet vil da anonymiseres. Det anonymiserte materialet vil da kunne oppbevares for videre forskningsarbeid.

Dine rettigheter

Så lenge du kan identifiseres i datamaterialet, har du rett til:

innsyn i hvilke personopplysninger som er registrert,

å få rettet personopplysninger

få slettet personopplysninger,

få utlevert en kopi av dine personopplysninger og

å sende klage til personvernombudet eller Datatilsynet om behandlingen av personopplysninger.

Hva gir oss rett til å behandle personopplysninger om deg?

Vi behandler opplysninger om deg basert på ditt samtykke.

På oppdrag fra NTNU har NSD – Norsk senter for forskningsdata AS vurdert at behandlingen av personopplysninger i dette prosjektet er i samsvar med personvernregelverket.

Hvor kan jeg finne ut mer?

Hvis du har spørsmål til studien, eller ønsker å benytte deg av dine rettigheter, ta kontakt med:

Gunnhild Saksvik-Raanes

Doktorgradsstipendiat

Institutt for lærerutdanning

NTNU

7491, Trondheim, Norway

Tlf: 41621593

Veileder for prosjektet: Trygve Solstad

Vårt personvernombud: Thomas Helgesen ved NTNU

NSD – Norsk senter for forskningsdata AS, på epost (personverntjenester@nsd.no) eller telefon: 55 58 21 17.

Med vennlig hilsen

Gunnhild Saksvik-Raanes

Prosjektansvarlig

Samtykkeerklæring

Jeg har mottatt og forstått informasjon om prosjektet digitale verktøy og tallforståelse, og har fått anledning til å stille spørsmål. Jeg samtykker til:

å delta i en uformell samtale om mine erfaringer knyttet til arbeidet med tallforståelse på 1. trinn.

at samtalen vil bli tatt opp på lydfil og senere transkribert og anonymisert.

Jeg samtykker til at mine opplysninger behandles frem til prosjektet er avsluttet, ca. mai 2023

(Signert av prosjektdeltaker, dato)

Appendix C: Approval NSD/Sikt

Sikt

Meldeskjema / Digitale verktøy og tallforståelse / Vurdering

Vurdering av behandling av personopplysninger

Referansenummer	Vurderingstype	Dato
329431	Standard	24.09.2019

Prosjekttittel
Digitale verktøy og tallforståelse

Behandlingsansvarlig institusjon
Norges teknisk-naturvitenskapelige universitet / Fakultet for samfunns- og utdanningsvitenskap (SU) / Institutt for lærerutdanning

Prosjektansvarlig
Gunnhild Saksvik-Raanes

Prosjektperiode
27.09.2019 - 08.04.2023

Kategorier personopplysninger
Alminnelige

Lovlig grunnlag
Samtykke (Personvernforordningen art. 6 nr. 1 bokstav a)

Behandlingen av personopplysningene er lovlig så fremt den gjennomføres som oppgitt i meldeskjemaet. Det lovlige grunnlaget gjelder til 08.04.2023.

Meldeskjema

Kommentar
Det er vår vurdering at behandlingen av personopplysninger i prosjektet vil være i samsvar med personvernlovgivningen så fremt den gjennomføres i tråd med det som er dokumentert i meldeskjemaet med vedlegg den 24.09.2019, samt i meldingsdialogen mellom innmelder og NSD. Behandlingen kan starte.

MELD VESENTLIGE ENDRINGER
Dersom det skjer vesentlige endringer i behandlingen av personopplysninger, kan det være nødvendig å melde dette til NSD ved å oppdatere meldeskjemaet. Før du melder inn en endring, oppfordrer vi deg til å lese om hvilke type endringer det er nødvendig å melde:

https://nsd.no/personvernombud/meld_prosjekt/meld_endringer.html

Du må vente på svar fra NSD før endringen gjennomføres.

TYPE OPPLYSNINGER OG VARIGHET
Prosjektet vil behandle alminnelige kategorier av personopplysninger frem til 08.04.2023.

LOVLIG GRUNNLAG
Prosjektet vil innhente samtykke fra de registrerte til behandlingen av personopplysninger. Vår vurdering er at prosjektet legger opp til et samtykke i samsvar med kravene i art. 4 og 7, ved at det er en frivillig, spesifikk, informert og utvetydig bekreftelse som kan dokumenteres, og som den registrerte kan trekke tilbake. Lovlig grunnlag for behandlingen vil dermed være den registrertes samtykke, jf. personvernforordningen art. 6 nr. 1 bokstav a.

PERSONVERNPRINSIPPER
NSD vurderer at den planlagte behandlingen av personopplysninger vil følge prinsippene i personvernforordningen om:

- lovlighet, rettferdighet og åpenhet (art. 5.1 a), ved at de registrerte får tilfredsstillende informasjon om og samtykker til behandlingen
- formålsbegrensning (art. 5.1 b), ved at personopplysninger samles inn for spesifikke, uttrykkelig angitte og berettigede formål, og ikke behandles til nye, uforenlige formål
- dataminimering (art. 5.1 c), ved at det kun behandles opplysninger som er adekvate, relevante og nødvendige for formålet med prosjektet
- lagringsbegrensning (art. 5.1 e), ved at personopplysningene ikke lagres lengre enn nødvendig for å oppfylle formålet

DE REGISTRERTES RETTIGHETER
Så lenge de registrerte kan identifiseres i datamaterialet vil de ha følgende rettigheter: åpenhet (art. 12), informasjon (art. 13), innsyn (art. 15), retting (art. 16), sletting (art. 17), begrensning (art. 18), underretning (art. 19), dataportabilitet (art. 20).

NSD vurderer at informasjonen om behandlingen som de registrerte vil motta oppfyller lovens krav til form og innhold, jf. art. 12.1 og art. 13.

Vi minner om at hvis en registrert tar kontakt om sine rettigheter, har behandlingsansvarlig institusjon plikt til å svare innen en måned.

FØLG DIN INSTITUSJONS RETNINGSLINJER
NSD legger til grunn at behandlingen oppfyller kravene i personvernforordningen om riktighet (art. 5.1 d), integritet og konfidensialitet (art. 5.1. f) og sikkerhet (art. 32).

For å forsikre dere om at kravene oppfylles, må dere følge interne retningslinjer og/eller rådføre dere med behandlingsansvarlig institusjon.

OPPFØLGING AV PROSJEKTET
NSD vil følge opp underveis (hvert annet år) og ved planlagt avslutning for å avklare om behandlingen av personopplysningene er avsluttet/pågår i tråd med den behandlingen som er dokumentert.

Lykke til med prosjektet!

Kontaktperson hos NSD: Karin Lillevold
Tlf. Personverntjenester: 55 58 21 17 (tast 1)

Appendix D: Interview guide teachers

1. **Introduction**
   I will start the interviews with a short introduction about the project, ethical considerations, and conditions regarding the recording of the conversation. Each of the teachers will need to sign the consent form before we start the conversation. The interview will be conducted at the teacher's workspace after they finish their teaching duties.

2. **Questions related to children's number sense and teachers' expectations.**

   - From your experience, could you describe what children know about numbers when they start school?
   - What number sense activities do you usually include in your teaching in the first semester of first grade?
   - What should the children know regarding number sense at the end of first grade?

   After these initial questions, the teachers are introduced to results from a previous study where we found that children in Early years education demonstrated a number sense knowledge that exceeded the learning goals for the second grade in the Norwegian curriculum.

   - Do these results surprise you?
   - How would you consider these results compared to what you discussed previously?

3. **The teachers are shown an overview of their child's results and asked to discuss the following questions:**

   - Can you discuss what these results tell about your child's number sense?
   - Does this align with what you have seen from the children in the class? Is there anything that surprises you?
   - What do these results tell you about this child's number sense?
   - Can you say something about how you would plan for the further learning process for this child?

4. **Questions related to previous assessments and further work:**

   - What is your experience with using tests to describe children's competence?

- In general, how well do you think tests (like the National numeracy test) describe your children’s mathematical knowledge?
- What should a teacher consider when using tests to describe children’s knowledge?
- More specific about items? Reliability?

Paper I

Developing a formative, digital tool to assess children's number sense when starting school

Gunnhild Saksvik-Raanes and Trygve Solstad

Norwegian University of Science and Technology, Norway; gunnhild.b.saksvik@ntnu.no, trygve.solstad@ntnu.no

We investigate how digital technologies can enrich teachers' formative assessment of number sense by describing the development process of a digital assessment tool for children starting school (five- and six-year-olds). Studying different aspects of validity, we focus on scoring and digital affordances. The quantitative analysis of the preliminary data from 101 children evidences the technical validity of the tool. We find that interactive assessment items add to the content validity of the tool. The interactive items provide qualitative data about students' number sense which cannot be captured by quantitative measures. At the same time, children may have greater difficulty interpreting more complex items. Our results support the view that a digital tool can be a useful supplement to the assessment of number sense. Further developments and approaches to investigating additional aspects of validity are discussed.

Keywords: Number sense, formative assessment, educational technology, elementary school mathematics.

Introduction

Children start school with considerable knowledge related to learning mathematics (Clarke et al., 2006). Assessing each child's number competence enables the teacher to plan effective and engaging teaching, but assessment can also be a challenging and time-consuming task. The rapid advance of digital technologies brings new opportunities for assessment in mathematics education. Developing digital formative assessment instruments has been highlighted as an area to prioritise in early mathematics education research (Ginsburg & Pappas, 2016). Specifically, to inform teachers about children's Foundational Number Sense (FoNS), Sayers et al. (2016) point out the need to "develop a diagnostic tool for teachers to assess individual grade one children's FoNS-related understanding" (p. 389).

We address this need by developing and analysing a digital assessment tool. The purpose of the tool is to inform teachers about the mathematical knowledge that children have already acquired when they start school. In a classroom setting, a digital tool can be an efficient supplement in the initial assessment process, especially since time typically does not permit one-to-one assessment interviews. In this paper, we focus on the opportunities and challenges associated with technology supplementing the teacher's formative assessment of students' number sense. We discuss how interactive tasks and records of qualitative data about the children's solutions and solution processes can strengthen the evidence for the construct validity of a digital assessment tool. Based on preliminary data from an on-going study, we address the following research question:

How do digital tasks support aspects of validity in the assessment of first-graders' number sense?

Frameworks

The FoNS model for operationalising number sense

To operationalise number sense we build on the FoNS model which describes the number-related skills that require instruction (Andrews & Sayers, 2015). The FoNS model provides a multi-layered, flexible, and relational definition of the number sense concept with eight interrelated categories: Number identification (NI), systematic counting (SC), number-quantity relationships (NQ), quantity discrimination (QD), representing numbers (RN), estimation (ES), simple arithmetic competence (AC) and awareness of number patterns (NP).

Assessment validity

In our context, assessment validity concerns the extent to which theory supports inferences made from test scores and how the evidence supports interpretations (Wolfe & Smith, 2007a). Evidence for validity can be found by looking at eight different aspects of validity: 1) content, 2) substantive, 3) structural, 4) generalisability, 5) external, 6) consequential, 7) interpretability, and 8) responsiveness (Wolfe & Smith, 2007b). In this paper, we focus on the content, substantive, and interpretability aspects of validity. The content aspect of validity addresses the relevance, representativeness, and the technical quality of the items. Documenting the purpose of the tool and the development process with expert reviews are part of the evidence that relates to the content aspect of validity. The substantive aspect of validity relies on the theoretical model we based the item development on. Here, we look at how the different items reflect children's overall number sense as described in the FoNS model. We focus on how the digital format can enhance content and substantive validity. Finally, we address how the number sense score can be interpreted by considering features that affect item difficulty.

Formative assessment of young children's knowledge

Formative assessment can be considered both an instrument and a process (Bennett, 2011). In this context, formative assessment refers to how the results of the assessment process are used to promote further learning (Black & Wiliam, 2018). Therefore, in addition to informing teachers about the children's present competence, a formative assessment tool should also support teachers in adjusting their teaching to meet the children's needs. Previous research has examined how task-based, one-to-one assessment interviews can provide crucial information to help teachers facilitate students' learning (Clarke et al., 2011). The digital medium can provide teachers with some of the features of assessment interviews, such as screen recordings of the students' solution processes on interactive items, as well as more traditional skill-based assessments characteristic of pencil-and-paper tests. Certain researchers highlight the transformative improvements that software can have on early mathematics education, both for helping children learn mathematics, and for providing guidance to teachers (Ginsburg, 2016). Hence, a digital assessment tool can guide instruction and improve students' opportunities to learn mathematics.

Methods

Procedure

About 50 schools in and around Trondheim municipality in Norway were invited to participate in the project, out of which eight of the interested schools were chosen. In this paper, we present preliminary data from 101 of the first-grade children (five to six years old) who participated in the study. A researcher visited the schools over a period of two months at the beginning of the school year. Groups of six to eight children carried out the assessment on separate tablet computers. The participants were seated so as not to get disturbed by each other's screens or sounds. All children were given the same instructions before they started the assessment and were free to finish at any time. Pre-recorded voice instructions were given for each item. For technical reasons, the assessment was presented in three separate units with different FoNS categories and increasing difficulty. Each child could decide whether to continue to the next unit or not. Most children completed the first two units, some completed all three units, and a few children completed only the first unit. There was no time limit for the items, but time on task was recorded for each item. The children typically spent between 15 and 25 minutes on the assessment in total.

Analytical procedures

Rasch measurement was used for quantitative analysis of the children's responses, using the Winsteps software. The Rasch model is a probabilistic measurement model that provides interval-scale measures of item difficulty and person skill on the same measurement scale in the unit of logits (Wright, 1977). On a basic level, Rasch analysis involves calculating the probability that a person with competence B answers correctly on an item with difficulty D. A person with higher competence always has a higher probability of successfully answering any item than a person with lower competence. An item that is more difficult, always has a lower probability of being successfully answered than an item that is less difficult, regardless of person ability.

All items were scored dichotomously, meaning that the children received one point for a correct answer and zero points for a wrong answer. In addition, to investigate and illustrate the potential of interactive items for enriching digital assessment, observations and screen recordings of the children's solutions underwent qualitative analysis.

Results and discussion

Development of the assessment tool: from framework to data collection

We developed items for each of the eight FoNS categories. The items were adapted from different cognitive and educational studies on number sense, standardised number sense tests, and formative assessment instruments (Ginsburg & Pappas, 2016; Davison et al., 2012).

A selection of items was presented to expert groups for them to adjust and determine the items that were the most suitable for measuring the number-sense construct. The expert reviews were performed by researchers in mathematics education familiar with the number-sense concept, target population and instrument development. Based on the first reviews, items from five number sense categories were selected for the first pilot study: Number identification, systematic counting, number and quantity, quantity discrimination, and arithmetic competence. Several pilot studies were conducted to refine the content and selection of items. The qualitative observations obtained from the first pilot study led to further changes, predominantly associated with technical issues and voice instructions. Certain items that were often misunderstood by the children were removed. The first pilot study was carried out at the end of the academic year, whereas the assessment was designed for children who were almost a year younger. To provide better estimates of the difficulty parameters, we conducted a second pilot study in a preschool right before the end of the academic year. This second pilot study led to further adjustments. The level of difficulty had to be adjusted to enable the overview of the number sense competence of all first graders. Subitising had also been included as a separate category at this point since conceptual subitising has been highlighted as a central aspect of number sense (Sayers et al., 2016). After analysis in expert groups, the decision was made to include the last three FoNS categories in the assessment tool: Estimation, number patterns, and representing number. The items from these categories were piloted in a fourth pilot study, along with the rest of the assessment, before commencing the main data collection in the fall of 2020. An overview of the final 78 items included in the tool is presented in Table 1. Two items did not contribute to the number sense measure as intended and were removed from the present analysis. We based the subsequent analysis on the remaining 76 items. The items varied from typical skill-based items, asking the child to identify a certain number symbol or quantity, to items capturing the child's solution process (later referred to as process items). The process items were designed to exploit the digital medium's potential for capturing some of the characteristic aspects of one-to-one assessment interviews. Of the 78 items, 10 were process items, and two of them will be discussed in more detail.

Scale	**Content**	**N(SB)**	**N(P)**
Number identification (NI)	Recognise numeral and meaning.	8	0
Systematic counting (SC)	Ordinality. Count to twenty and back from an arbitrary digit.	6	2
Number and quantity (NQ)	Cardinality. One-to-one correspondence between symbol and quantity.	6	4
Quantity discrimination (QD)	Compare quantities. Vocabulary: larger, smaller, more than, less than.	6	2
Representing number (RN)	Different representations of numbers and part - whole aspects.	4	1
Estimation (ES)	Estimate the size of a set and position on a number line.	6	0
Arithmetic competence (AC)	Operate on small sets by using addition or subtraction	14	1
Number patterns (NP)	Continue or complete a number sequence.	7	0
Subitising (SU)	Perceive quantity without counting. Perceptual and conceptual. Timed.	11	0

Table 1. Overview of the number sense items within each FoNS category. N(SB): Number of skill-based items. N(P): number of process items

Development of a specific item involving children's solution strategies

NQ10 was developed within the category of number and quantity (NQ) to assess the ability to count a number of objects and identify the number symbol that represents that amount. The item was adapted from similar items used in interview settings (Malofeeva et al., 2004). We present the first version of the item in Figure 1A. Expert groups discussed the specific number of objects that would enable insights into the child's structuring of quantities, how to give the most succinct and accurate voice instructions, and how the objects could be placed to encourage the child to manipulate them.

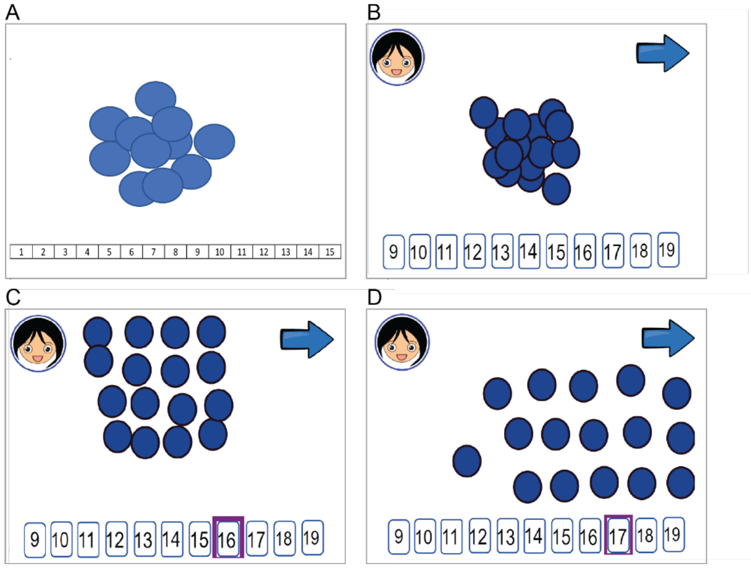


Figure 1. Development and responses to Item NQ10
A. First version; B. Final version; C. Final response of Child 1; D. Final response of Child 2

In the final version of item NQ10 (Figure 1B) the children are asked to arrange the objects in a manner that is easy to count. The software provides both a recording of the motion and the final positions, which allowed us to study how the child rearranges the objects to simplify the counting process.

In the given examples, we see how two first-graders structured the objects. Child 1 structured the 16 objects into four groups of four and selected the correct answer. Child 2 structured the objects into three groups of five, placing one object separately, and then selected 17 as the answer. If not just a random mistake, the error might be related to the number symbols (rather than the quantity involved), the aspects of cardinality, or the structure of odd and even numbers. Such access to the qualitative aspects of the responses can give the teacher valuable insight into each child's solution process and more information about the child's understanding than simply knowing whether the response was right or wrong. Therefore, these process items can provide qualitative data to support the assessment process and contribute to the evidence for both the content and substantive aspects of validity. Process items cannot replace the insights gained through human interaction in one-to-one assessment interviews. However, the qualitative recordings can support the validation of the assessment and be valuable for screening, individual follow-up, and teaching purposes.

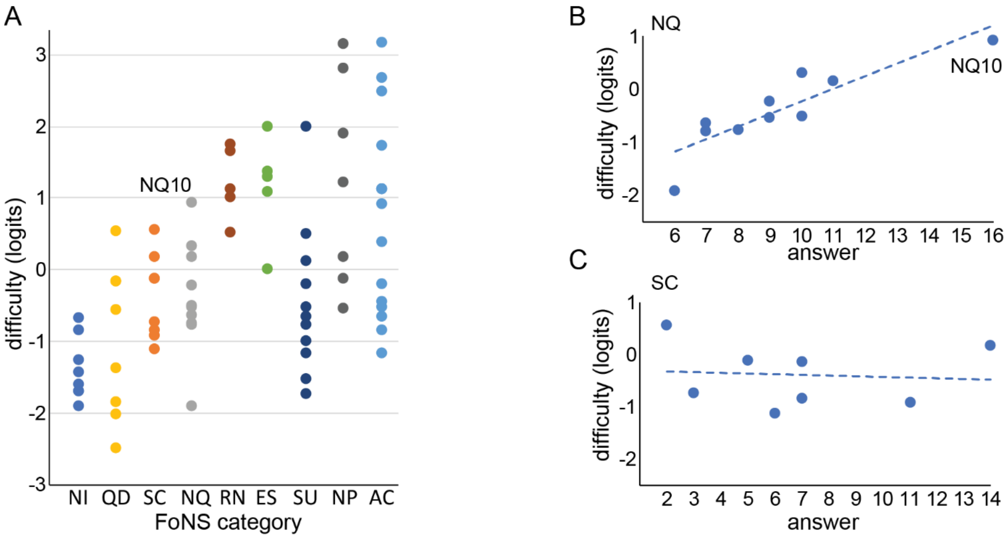


Figure 2. Item difficulty. A. Plotted as a function of the FoNS categories arranged by maximum item difficulty. B, C. Plotted as a function of the correct answer to the item in the categories number and quantity (NQ) and systematic counting (SC). The dashed line is the linear regression line. Item NQ10 is indicated in the plots (difficulty = 0.93; answer = 16). The abbreviations are defined in Table 1

Test quality: technical aspects of construct validity

The person reliability value of the 76 items analysed in this paper was 0.88 (corresponding to a Cronbach's alpha of 0.91), which is typically considered productive for measurement.

As a whole, the items were well targeted to the group of school-starters and were slightly easy, having the value of 0 ± 1.31 logits (mean ± sd), while the children scored 0.51 ± 0.93 logits (mean ± sd). The larger standard deviation of the items indicates that only a few children scored at the top and bottom of the measurement scale. This gives the assessment tool the necessary range to track the children's number sense as it develops during their first year of school, for example, by comparing the assessments from the first and second half of the school year.

If a single aggregate score is used to measure number sense, individual items need to measure the same number sense construct. Evidence for this can be found in the infit and outfit measures of the items. An item with fit values close to 1 is considered to measure the same construct as the rest of the items. For this assessment tool, most items fit well with the Rasch model, with a mean item infit mnsq of 0.98 ± 0.15 (mean ± sd) and mean item outfit mnsq of 1.02 ± 0.42.

The group of process items also measured the same construct as the other items. One potential exception was item RN4, with infit mnsq 1.21 and outfit mnsq of 2.2, which is high. Excluding RN4, the process items had a mean infit mnsq of 1.02 ± 0.11 and mean outfit mnsq of 0.99 ± 0.17.

Item RN4, in which the child uses dice to compose the quantity ”four” in three different ways, was adapted from a composing number task used in interview settings (Clements et al., 2008). Formulating clear instructions for such complex tasks was a general challenge of the item development and might be related to the misfit of this item. Representing numbers seems to be different from other number sense categories, as three out of five items from this category had infit values greater than 1.2 (one process and two skill-based items). One reason for items from this category to stand out as more misfitting than items form other categories

might be the complexity of this domain. Representing numbers in different ways is a rich domain which includes several other categories of number sense, such as arithmetic competence and connections between number and quantity (Andrews and Sayers, 2015).

How can the number sense score be interpreted? Here, we give two examples supporting the interpretability aspect of validity of the assessment. First, different FoNS categories had different ranges of difficulty (Figure 2A). This means that a child's aggregated measure is indicative of whether that child can solve the tasks from each category. Second, for some FoNS categories, such as number and quantity, the difficulty correlated with the numerical value of the answer (Figure 2B). The relation between numerical value and item difficulty suggests that the aggregate measure is predictive of the range of numbers the child can process confidently. For other categories, such as systematic counting, the difficulty was not clearly related to the numerical value of the answer (Figure 2C), indicating that task difficulty was largely determined by the content of the tasks in some categories.

Conclusions

Presenting parts of an on-going study, we have described the development of a tool to assess first grader's number sense. The digital assessment tool is based on the FoNS model (Andrews & Sayers, 2015) and has been subjected to expert reviews. The use of interactivity provides opportunities for simulating aspects of interview tasks, which adds to the content aspect of the validity of the assessment. The Rasch analysis of the first graders' responses indicates that the technical quality of the assessment tool is high. Taken together, these results indicate that a digital assessment tool has the potential to provide teachers with information about their students' number sense.

We argue that the digital format can supplement the teachers' informal assessments and offer a valid, reliable, and more complete alternative to paper-and-pencil assessment by incorporating aspects of one-to-one assessment interviews. The use of interactive process items gives us the opportunity to adapt tasks that were previously reserved for one-to-one assessments to a setting in which teachers can gain information about their students' number sense in a less time-consuming manner. It remains to investigate how the digital assessment may affect test conditions and whether the level of engagement in the digital assessment is different from that of comparable written assessments.

An important validity aspect of a formative assessment tool is how the tool is used to improve children's learning. Describing formative assessment as a process and validating a tool to be used in this process, entails presenting the scores in a manner that is useful for the teacher. In a classroom setting, the teacher's informal assessments are often based on several different interpretations, which leads to a measurement issue in the formative assessment (Bennett, 2011). The integration of fundamental measurement principles with digital technology may help ameliorate this issue. To be able to "develop curriculum support tools for teachers to plan an explicit incorporation of FoNS categories in their teaching" (Sayers et al., 2016, p. 389), the other aspects of the validity of the tool must be considered. As a next step, we need to assess the consequential validity of the claim that the tool can usefully supplement the teacher's formative assessment of number sense.

Acknowledgements

We would like to thank everyone involved in this project for their contribution to aspects such as technical implementation, design, expert reviews, data collection and analysis.

References

Andrews, P. & Sayers, J. (2015). Identifying Opportunities for Grade One Children to Acquire Foundational Number Sense: Developing a Framework for Cross Cultural Classroom Analyses. *Early Childhood Education Journal*, *43*(4), 257–267. https://doi.org/10.1007/s10643-014-0653-6

Bennett, R. E. (2011). Formative assessment: A critical review. *Assessment in Education: Principles, Policy & Practice*, *18*(1), 5–25. https://doi.org/10.1080/0969594X.2010.513678

Black, P. & Wiliam, D. (2018). Classroom assessment and pedagogy. *Assessment in Education: Principles, Policy & Practice*, *25*(6), 551–575. https://doi.org/10.1080/0969594X.2018.1441807

Clarke, B., Cheeseman, J. & Clarke, D. (2006). The mathematical knowledge and understanding young children bring to school. *Mathematics Education Research Journal*, *18*(1), 78–102. https://doi.org/10.1007/BF03217430

Clarke, D., Clarke, B. & Roche, A. (2011). Building teachers' expertise in understanding, assessing and developing children's mathematical thinking: The power of task-based, one-to-one assessment interviews. *ZDM*, *43*(6), 901–913. https://doi.org/10.1007/s11858-011-0345-2

Ginsburg, H. (2016). Helping early childhood educators to understand and assess young children's mathematical minds. *ZDM*, *48*(7), 941–946. https://doi.org/10.1007/s11858-016-0807-7

Ginsburg, H. & Pappas, S. (2016). Invitation to the birthday party: Rationale and description. *ZDM*, *48*(7), 947–960. https://doi.org/10.1007/s11858-016-0818-4

Malofeeva, E., Day, J., Saco, X., Young, L. & Ciancio, D. (2004). Construction and Evaluation of a Number Sense Test With Head Start Children. *Journal of Educational Psychology*, *96*(4), 648–659. https://doi.org/10.1037/0022-0663.96.4.648

Sayers, J., Andrews, P. & Boistrup, L. B. (2016). The role of conceptual subitising in the development of foundational number sense. In W. A. Meaney T., Helenius O., Johansson M., Lange T. (Eds.), *Mathematics education in the early years* (pp. 371–394). Springer. https://doi.org/https://doi.org/10.1007/978-3-319-23935-4_21

Wolfe, E. W. & Smith, E. V. (2007a). Instrument development tools and activities for measure validation using Rasch models: Part I--instrument development tools. *Journal of Applied Measurement*, *8*(1), 97–123.

Wolfe, E. W. & Smith, J. E. V. (2007b). Instrument development tools and activities for measure validation using Rasch models: Part II--validation activities. *Journal of Applied Measurement*, *8*(2), 204–234.

Wright, B. D. (1977). Solving Measurement Problems with the Rasch Model. *Journal of Educational Measurement*, *14*(2), 97–116. http://www.jstor.org/stable/1434010

Paper II

Factors that influence the difficulty level of digital arithmetic assessment items for first-grade students

Gunnhild Saksvik-Raanes, Trygve Solstad

Factors that influence the difficulty level of digital arithmetic assessment items for first-grade students

Gunnhild Saksvik-Raanes[1] and Trygve Solstad[1]

[1]Norwegian University of Science and Technology, Norway;
gunnhild.b.saksvik@ntnu.no, trygve.solstad@ntnu.no

Digital technologies enable new possibilities for the assessment of mathematical competence. When designing an assessment, it is essential to know how different design elements affect both the item difficulty and the strategies used by the children. In this paper, we investigate digital items that were designed to measure arithmetic competence as a component of the Foundational Number Sense (FoNS) framework for five- and six-year-old children. A Rasch analysis of the performance of 302 Norwegian children showed that the type of arithmetic problem and the magnitude of the answer strongly affected an item's difficulty level. Our qualitative observations indicated that certain additional design elements of the items might have influenced both the items' difficulty and the children's solution strategies. From a mixed methods perspective, we discuss the potential of different design elements to better assess children's understanding of numbers.

Keywords: Assessment, digital technology, numbers sense, arithmetic competence, primary school.

Introduction

Digital technologies bring both constraints and affordances to assessment in mathematics education (Threlfall et al., 2007). When assessing young children's mathematical competence, using a digital medium can alleviate the effect of irrelevant demands, such as reading or writing skills. At the same time, we might add elements in the design process of a digital item that could affect the assessment in unintended ways. Carefully designing digital assessment items might enable us to improve assessments and tell us more about the children's solution processes (Saksvik-Raanes & Solstad, 2021). To realise the full potential of digital assessments, we need to know more about how children perceive the different design elements of digital items and what strategies they use to solve them.

In this paper, we investigate digital arithmetic items for five- and six-year-olds from a mixed methods perspective and pose the following research question: Which design elements influence the level of difficulty of digital arithmetic assessment items, and how do the design elements influence the strategies that first-grade children use to solve these items?

Frameworks

Arithmetic competence as a part of the FoNS model

The Foundational Number Sense (FoNS) model describes the number-related skills that require instruction (Andrews & Sayers, 2015). In FoNS, the number sense concept is defined as multi-layered, flexible and relational. The FoNS model divides the number-related skills into eight interrelated categories: number identification, systematic counting, number–quantity relationships, quantity discrimination, representing numbers, estimation, simple arithmetic competence, and awareness of number patterns. In this paper, we focus on simple arithmetic competence, which is described as a child's ability to manipulate small sets through addition or subtraction.

Item design

The arithmetic items presented in this paper were designed based on four main categories: change and combine (sum items) as well as compare and equalise (difference items) (Carpenter & Moser, 1984). Previous studies have shown that children in kindergarten are highly capable of solving such items through modelling the situations in the problem (Carpenter et al., 1993). Sum items are composed in two ways. Either with one initial quantity and an action that causes a change (join), or with two initial quantities that may be considered separately or as a part of a whole. Difference items involve comparing two quantities to determine the difference between them. Equalise problems include an additional action that is to be performed to make the two sets equal.

In addition to problem type, the difficulty of the items was expected to depend on three further design elements. Some items used pictorial representations of numbers, and other items used symbolic representations of numbers. The items involved different numerical values between one and twenty. We balanced the number of items involving small (< 10) and large (≥ 10) numbers and items having ordered and unordered response buttons (see Figure 3).

Methods

Participants and procedure

Fifteen arithmetic items were solved by 302 first-grade children who were a part of a larger study that investigated 368 children's number sense using digital assessment tools. To select participants for the project, we invited about 50 elementary schools in and around Trondheim municipality in Norway to participate in the project. Eight of the interested schools were chosen to participate and all the 1st grade children in these schools carried out the assessment. The children were five and six years old.

A researcher visited the schools over a period of two months at the beginning of the school year. Groups of six to eight children carried out the assessment on separate tablet computers. The participants were seated in such a manner that they would not be disturbed by each other's screens or sounds. All children were given the same instructions before they started the assessment and were free to finish it at any time. Pre-recorded voice instructions were given for each item. The arithmetic items appeared at the end of the full assessment. There was no time limit for the items, but the time taken for each item was recorded. The children typically spent between 15 and 25 minutes on the full assessment, of which about one-fifth comprised arithmetic items.

Qualitative data from individual interviews conducted with 19 first- grade children solving the arithmetic items, were collected independently as a part of a master's degree project (Schjølberg, 2021). The goal of the interviews included in the master's project was to get an overview of the children's strategies. One of the strategies applied by one of the students who participated in the master's project is included in this paper. The interviews were carried out at about the same time as the main data collection.

All the described studies have been approved by the Norwegian Centre for Research Data, and the necessary guidelines related to depersonalisation and parental consent have been followed.

Items

The 15 arithmetic items were designed to investigate the different aspects of the children's arithmetic competence that could influence item difficulty. Four items involved the difference between two numbers. Two of these 'difference items' were compare problems, and two were equalise problems (Carpenter & Moser, 1984). The compare items involved small numbers (< 10), while the equalise problems involved large numbers (≥ 10).

Eleven items asked for the sum of two numbers. Eight of these 'sum items' included the systematic variation of three design elements: (i) small (< 10) or large (≥ 10) answer, (ii) symbolic or pictorial representation of the problem and (iii) ordered or unordered response buttons (see Figure 3).

A priori, we expected the difference items to be more difficult than the sum items, the items with large numbers to be more difficult than those with small numbers, the items with symbolic representations of numbers to be more difficult than those with pictorial representations and the items with unordered response buttons to be more difficult than those with ordered response buttons.

Analysis

All items were scored dichotomously, meaning that the children received one point for a correct answer and zero points for a wrong answer. Rasch measurement was used for the quantitative analysis of the children's responses using the Winsteps software (Linacre, 2017). The Rasch model is a probabilistic measurement model that provides interval-scale measures of item difficulty and person skill on the same measurement scale in units of logits (Wright, 1977). The probability that person v scores 1 point on item i depends on the difference between the skill of person v, β_v and the difficulty of item i, δ_i according to

$$\boldsymbol{P}\ \{\boldsymbol{X_{vi}} = \mathbf{1}|\boldsymbol{\beta_v}, \boldsymbol{\delta_i}\} = \frac{\boldsymbol{e}^{(\boldsymbol{\beta v} - \boldsymbol{\delta i})}}{\mathbf{1} + \boldsymbol{e}^{(\boldsymbol{\beta v} - \boldsymbol{\delta i})}}$$

Winsteps implements the joint maximum likelihood estimation (JMLE) algorithm to estimate the parameters of this model.

The excerpt from the individual interviews demonstrates how some children used the available resources on the screen to find the right answer to the problems. The qualitative data was analysed using a thematic analysis (Bryman, 2016).

Results and discussion

Task type

From Figure 1, we see that the type of task strongly influenced the difficulty of the items. As expected, the four difference items were also the four most difficult arithmetic items (Figure 1, orange markers). An independent samples t-test between the four difference items and four comparable summation items (symbolic representations involving large and small numbers and ordered and unordered response buttons) showed that this difference was significant (p = 0.026; df = 6).

Surprisingly, within the difference category, both compare items had higher difficulty than the two equalise items. The compare items were more difficult despite involving small numbers, while the equalise items involved large numbers and came with more complex voice instructions.

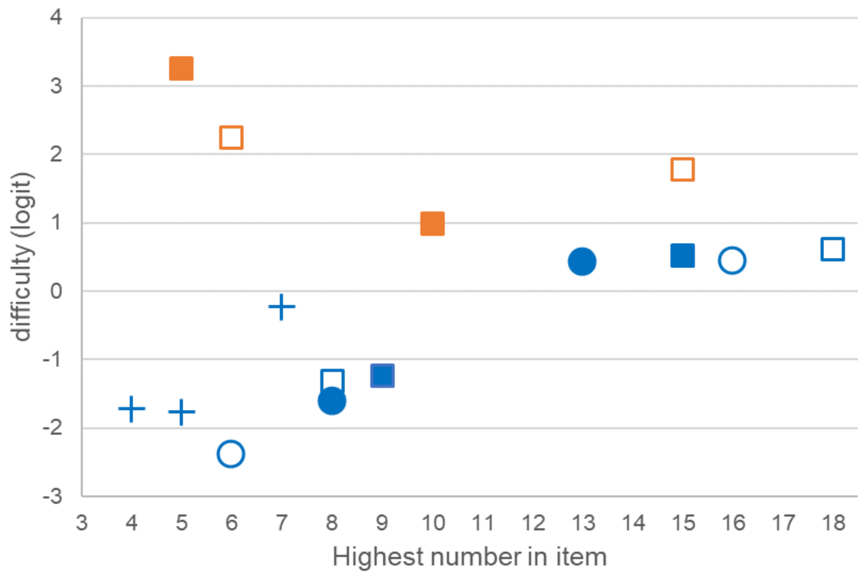


Figure 1: Item difficulty ordered by the highest number involved in the item

The items are categorised as (i) sum item (blue markers) and difference item (orange markers), (ii) pictorial representation (circular markers) and symbolic representation (square markers), and (iii) ordered response buttons (open markers), unordered response buttons (filled markers) and no response buttons (plus markers)

The compare and equalise items were visually identical (Figure 2), and three design elements differentiated them: the magnitude of the answer, the order of the response buttons and the voice instruction given. The answer was less than 10 for both compare items, while the answer was greater than 10 for both equalise items. The following voice instruction was given for the compare items: "How many more marbles are there in the blue box?". For the equalise problems, the following voice instruction was given: "There should be an equal number of marbles in each box. How many more should the red box have?".

In the design process, the difference items were challenging to create in a way that would enable all children to understand the given voice instructions. We wanted to keep the instructions as simple as possible to adapt to the attention span of the target group. At the same time, the compare and equalise problems represent two semantically different problems. The equalise problems involve one more step than the compare problems, as an action is performed on one of the two groups when comparing the quantities. We therefore expected that the equalise items would be more difficult. However, Figure 1 shows that the compare items were the most difficult.

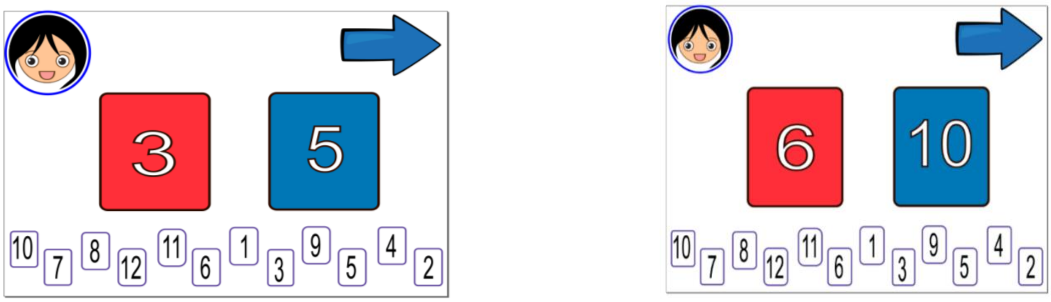


Figure 2: Two difference items

Left: item 13 (compare). Right: item 15 (equalise)

One reason why the compare items appear to be more difficult could be that the added action in the instructions for the equalise items make them more concrete, and this might have aided the children's comprehension of the items. Carpenter et al. (1993) found the kindergarteners in their study to be highly competent in solving compare problems through modelling. It is also possible that some children ignored the "more" word in the instructions for the compare tasks and interpreted it to mean "how many marbles are there in the blue box?". These results indicate that simplified instructions in word problems may lead to more misunderstandings and reduce the child's possibilities for modelling the situation.

The level of abstraction in the illustrations of these four items might also have contributed to their relatively high difficulty compared to the sum items. The children's previous experiences could also have played a role in determining the level of difficulty, as it seems that they were more familiar with the language-related problems that involved addition than subtraction.

Taken together, these results underline the importance of carefully investigating the various design elements when developing digital assessment items.

Magnitude of the answer

For the sum items, we found that difficulty was strongly correlated with the magnitude of the answer of an item. The Pearson correlation between difficulty and answer magnitude was $r = 0.88$ ($p < 0.001$) for all 11 sum items and $r = 0.96$ ($p < 0.001$) for the eight sum items that had a shared problem structure (Figure 1). In particular, the four sum items with a large answer were 2.1 logits more difficult than the corresponding four sum items with a small answer on average. An independent samples t-test showed that this difference was significant ($p < 0.001$; df = 6).

Number representations

A pictorial representation of a number is often thought to be easier to understand than its more abstract, symbolic representation. However, in the group of the eight sum items that shared a problem structure, we found no significant difference in the difficulty between the four items with symbolic representations (blue squares in Figure 1) and the four corresponding items with pictorial representations (blue circles in Figure 1) ($p = 0.65$; $df = 6$; independent samples t-test). One reason for this might be that the response buttons were written in the symbolic representation. Thus, knowing the correct answer only verbally would not be sufficient to provide a correct response. Indeed, from the qualitative data, we observed that some children knew how to verbally count to 20 without recognising the corresponding written numerals (see the next section).

Order of the response buttons

Based on pilot studies, we had the a priori expectation that unordered response buttons would increase the difficulty of the items because they do not easily allow children to rely on verbal counting strategies. However, at least at first glance, the structure of the response buttons did not seem to strongly influence the difficulty of the items (Figure 1; open vs filled markers). An independent samples t-test between the four sum items with ordered response buttons and the four sum items with unordered response buttons was not significant ($p = 0.88$; $df = 6$).

On closer inspection, the four sum items with large answers were found to be of similar difficulty (Figure 1) even though the two items with ordered response buttons had larger answers than the two items with unordered response buttons. It is therefore possible that the ordered response buttons made the two tasks with the largest answers easier to solve. The latter interpretation is substantiated by qualitative analyses of the children's solution strategies. One example is item 10, which involved numbers that some children were not very familiar with. After the voice instruction "What is sixteen and two altogether?", the child was to choose the correct answer from the response buttons at the bottom of the screen. The qualitative observations gathered during the data collection led us to carry out a small qualitative interview study on the children's solution strategies.

Qualitative observations of Agnes's strategies

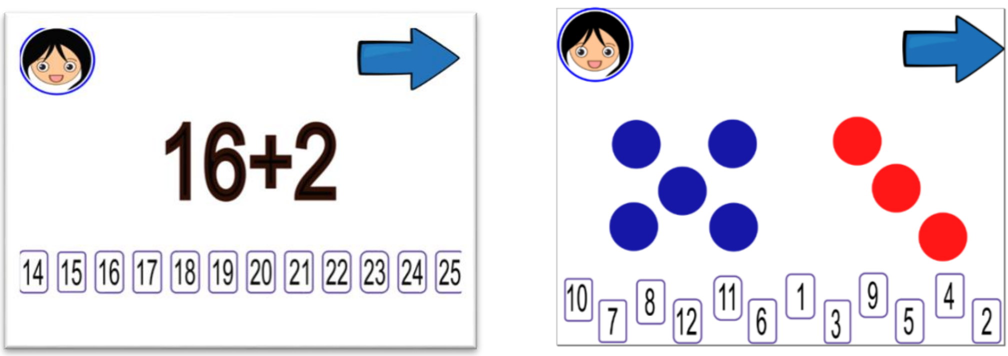


Figure 3: Sum items with systematic variation in the design elements

Left: item 10 with large answer, symbolic representations and ordered response buttons. Right: item 5 with small answer, pictorial representations, and unordered response buttons

In item 10, one of the children, Agnes, used the number alternatives on the screen to find the right answer, but she did not know what numeral she ended up with:

Researcher: What is 16 and 2 altogether?
(..)
Agnes: It is... 16 and 2…
(..)
Agnes: Wait… and then we go 1-2.
(Agnes points to 16 and makes two jumps with her finger on the numerals to the right)
(..)
Researcher: What are you thinking?
Agnes: That one.
(Points to 18)
Researcher: Do you know what number that is?
Agnes: 1-2-3-4-4-5-6-7-8-9-10-11-12-13-14-15-16-17-18. Eighteen!
(Counting to 14 and then pointing to the numerals on the screen)

Agnes used the buttons to find the numeral that displayed the answer when she was unable to recollect the numerals after 14. In the design process, we did not expect the children to use the number alternatives in this way. These observations also emphasise the importance of investigating the available resources and how these could affect children's solution strategies.

To obtain a more fine-grained analysis of the role of ordered and unordered response numerals, we need to investigate an instrument in which items with ordered and unordered response buttons are designed with identical arithmetic problems.

Conclusions

To investigate the factors that influence the level of difficulty in digital assessment items for arithmetic, we have looked at the role of problem type, representations, numerical values and differently ordered response buttons. We have also considered how children may use the ordered response buttons to find the correct answer for an item.

The strongest determinant of item difficulty was the type of problem: difference items were more difficult than sum items, and compare items were more difficult than equalise items. The second strongest determinant of item difficulty was the numerical value of the item's answer. Whether the problem was presented in a symbolic or pictorial form did not affect item difficulty. Finally, although we could not conclusively determine the influence of ordered or unordered response buttons, our data indicates that ordered response buttons allow children who have not yet acquired mastery over large numerals to use these buttons as a number sequence that helps them solve the problem. Including both kinds of response buttons might help distinguish between the children's knowledge of large numerals and their reasoning regarding the number sequence or with a number line.

While digital technologies continue to influence the assessment of students' mathematical competence with its new possibilities, it is also important to consider the technical and methodological challenges involved in this development (Nortvedt & Buchholtz, 2018). There are many aspects to consider when investigating the various elements that affect children's solution processes when interacting with digital technologies. To ensure the validity of such assessments, it is important that future research looks more into how the various possibilities that digital technologies enable can both improve and hinder students' performance. One way forward could be to compare items that have different design elements but similar answers. Looking more directly at a greater variety of both word problems and symbolic items can allow us to determine how the different contents affects the difficulty of the items. With the use of digital technology, we could also look more closely into young children's competence for solving digital word problems. For instance, one could record the children's solution process while they are introduced to a variety of word problems with more elaborate instructions and pictorial representations. The use of different digital aids, with a larger degree of interactivity, could enable us to study in more detail how the children model and use different strategies to solve the problems.

References

Andrews, P., & Sayers, J. (2015). Identifying opportunities for grade one children to acquire Foundational Number Sense: Developing a framework for cross cultural classroom analyses. *Early Childhood Education Journal*, *43*(4), 257–267. https://doi.org/10.1007/s10643-014-0653-6

Bryman, A. (2016). *Social research methods* (5th ed.). Oxford University Press.

Carpenter, T. P., Ansell, E., Franke, M. L., Fennema, E., & Weisbeck, L. (1993). Models of problem solving: A study of kindergarten children's problem-solving processes. *Journal for Research in Mathematics Education*, *24*(5), 428–441. https://doi.org/10.5951/jresematheduc.24.5.0428

Carpenter, T. P., & Moser, J. M. (1984). The acquisition of addition and subtraction concepts in grades one through three. *Journal for Research in Mathematics Education*, *15*(3), 179–202. https://doi.org/10.2307/748348

Dewey, J., Alexander, T. M., & Hickman, L. A. (1998). *The essential Dewey: 1: Pragmatism, education, democracy* (Vol. 1). Indiana University Press.

Linacre, J. M. (2017). *Winsteps® Rasch measurement computer program. User's Guide.* Beaverton, OR: https://www.winsteps.com/

Nortvedt, G. A., & Buchholtz, N. (2018). Assessment in mathematics education: Responding to issues regarding methodology, policy, and equity. *ZDM – Mathematics Education*, *50*(4), 555–570. https://doi.org/10.1007/s11858-018-0963-z

Saksvik-Raanes, G., & Solstad, T. (2022). Developing a formative, teacher-oriented, digital tool to assess number sense in school starters. *NORMA20 Proceedings*, Oslo Norway.

Schjølberg, A. (2021). *"Jeg må prøve å telle" En kvalitativ studie om elevers strategier i arbeid med tallforståelsesoppgaver på første trinn* (p. 87). [Master thesis] Norwegian University of Science and Technology.

Threlfall, J., Pool, P., Homer, M., & Swinnerton, B. (2007). Implicit aspects of paper and pencil mathematics assessment that come to light through the use of the computer. *Educational Studies in Mathematics*, *66*(3), 335. https://doi.org/10.1007/s10649-006-9078-5

Wright, B. D. (1977). Solving measurement problems with the Rasch model. *Journal of Educational Measurement*, *14*(2), 97–116. http://www.jstor.org/stable/1434010

Paper III

nordisk barnehageforskning
nordic early childhood educational research

Forskningsartikkel | Vol. 20, No. 2, 2023, s. 5–28 | ISSN: 1890-9167

Et innblikk i barns tallforståelse på slutten av barnehageårene gjennom digitale oppgaver

Gunnhild Saksvik-Raanes*, Trygve Solstad og Yvonne Grimeland
NTNU Norges teknisk-naturvitenskapelige universitet, Norge

***Korrespondanse:** Gunnhild Saksvik-Raanes, e-post: gunnhild.b.saksvik@ntnu.no

Sammendrag
I denne studien undersøkte vi tallforståelsen hos 77 femåringer fra fem ulike barnehager med digitale oppgaver. Kvantitative analyser av datamaterialet viste at det allerede før skolestart var stor spredning i barnas tallforståelse. Sammenlignet med det som er kjent fra før om norske barns utvikling av tallforståelse i barnehagealder, viste gruppen vi undersøkte overraskende godt utviklet tallforståelse. Vi diskuterer hva resultatene kan bety for den pedagogiske aktiviteten i barnehagen med tanke på å tilpasse aktiviteter til barns tallforståelse.

Nøkkelord: *FoNS; kartlegging; Rasch analyse; tallforståelse*

Abstract
An insight into five-year-old children's number sense in early childhood education using digital tasks
In this study we investigated the number sense of 77 five-year-old children from five early childhood education institutions using digital assessment tasks. From quantitative analyses of the data, we found a large variability in the children's number sense already before the start of formal education. Compared to what is known about Norwegian children's development of number sense in early childhood education, many children showed a surprisingly well-developed number sense. We discuss what these results could mean for learning activities in early childhood education.

Keywords: *assessment; FoNS; Rasch analysis; number sense*

Gjesteredaktører: Tamsin Meaney, Elin K. L. Reikerås og Camilla N. Justnes

Citation: Saksvik-Raanes, G., Solstad, T. & Grimeland, Y. (2023). Et innblikk i barns tallforståelse på slutten av barnehageårene gjennom digitale oppgaver. *Nordisk barnehageforskning*, *Special issue: Matematikk i barnehagen, 20*(2), 5–28. https://doi.org/10.23865/nbf.v20.279

Introduksjon

De siste års forskning i Norge har gitt større forståelse for kompleksiteten i det barnehagelærere faktisk gjør når de arbeider med matematikk i barnehagen (Justnes & Mosvold, 2021; Sæbbe & Mosvold, 2020). Videre har vi fått økt innsikt i hvordan barnehagebarn tilnærmer seg matematikken i lek og argumenterer gjennom flere modaliteter (Nergård, 2021), og hvordan muligheter for læring skapes mellom barn og barnehagelærer (Breive, 2020). Barnehagelærerens kompetanse kan ha innflytelse på barnets deltakelse og engasjement i den matematiske samhandlingen i barnehagen, hvor mye tid som brukes på matematiske aktiviteter, og nivået på den matematiske diskursen (Hundeland et al., 2020). En systematisk tilnærming til matematiske læringsaktiviteter og kompetanseheving i personalet kan også bidra til å styrke barnas matematiske utvikling (Rege et al., 2021).

En av utfordringene barnehagelærere møter er å tilpasse innholdet i matematiske aktiviteter til barnas eksisterende kunnskap (Rinvold & Erstad, 2015). Utfordringen er også kjent fra andre lands forskning hvor det blant annet har blitt påpekt at aktiviteter i barnehage og skole ofte undervurderer hva barn egentlig kan (Clarke et al., 2006; Engel et al., 2013). Mer kunnskap om hva vi kan forvente at barn forstår om tall vil gi oss muligheten til å legge bedre til rette for at alle barn i barnehagen får møte meningsfulle matematiske aktiviteter, uavhengig av personlig utgangspunkt og sosial eller geografisk tilhørighet. Slik kunnskap vil også kunne bidra til å styrke kontinuiteten i opplæringen fra barnehage til skole (Hogsnes & Moss, 2014; Sundtjønn et al., 2021).

Selv om verdien av den matematiske kunnskapen barn tar med seg inn i skolen er anerkjent internasjonalt (Clarke et al., 2006; Ginsburg, 2002), er det få studier som har undersøkt barns tallforståelse i norsk og nordisk kontekst, både før og like etter skolestart (Reikerås, 2016). I den nordiske barnehagekonteksten evalueres og vurderes først og fremst kvaliteten på utdanningssystemet og hva slags aktiviteter det blir tilrettelagt for. Vurdering av det enkelte barns forutsetninger og kunnskaper er lite vektlagt med mindre man er bekymret for barnets utvikling (Urban et al., 2022). Følgelig er de fleste eksisterende kartlegginger og studier av matematikkunnskapene til barn under sju år designet for å identifisere de barna som trenger ekstra oppfølging. Denne typen kartlegginger inneholder først og fremst oppgaver det er forventet at flertallet av barna får til å løse, og de er derfor ikke egnet til å si noe om hva eller hvor mye det er *typisk* at barn i denne aldersgruppen får til (Reikerås, 2016). Også mange internasjonale kartlegginger er bundet av en slik tak-effekt (Benoit et al., 2013; Mix et al., 2014). Dermed har vi begrenset informasjon om tallforståelsen til flertallet av norske barn, og få verktøy for å tilegne oss slik informasjon innenfor rimelige ressursrammer.

Det kan være utfordrende få innsikt i barns matematiske kunnskaper. Skriftlige kartlegginger er effektive (Lopez-Pedersen et al., 2021), men validiteten og egnetheten til disse verktøyene er begrenset av faktorer som kontekst og språk, barns begrensede lese- og skriveferdigheter, samt negative opplevelser forbundet med tidsbegrensede vurderingsoppgaver. Individuelle kartleggingsintervju (Clarke et al., 2011) og observasjoner av hvordan barnet naturlig anvender kunnskaper og ferdigheter i lek og hverdagslige aktiviteter (Bergsmo et al., 2020; Wager & Parks, 2016), kan derfor være å foretrekke for å få et best mulig bilde av hvert enkelt barns evne til matematisk tenking, handling og kommunikasjon. Slike intervjuer er på sin side tidkrevende, forutsetter en bestemt kompetanse hos hver enkelt kartlegger, og favner ofte et smalere spekter av oppgaver eller barn. Digitale plattformer gir mulighet til å balansere noen av egenskapene til de ulike kartleggingsmetodene og er blitt anerkjent som nyttige verktøy for å få innsikt i barns matematiske kunnskaper og videre behov (Clements et al., 2021; Ginsburg & Pappas, 2016).

Med et økt fokus på digitale verktøy i det matematiske læringsarbeidet i barnehagen (Carlsen et al., 2016; Lembrér & Meaney, 2016) følger det også muligheter til større innsikt i barns matematiske kunnskaper. En slik mulighet er at man kan lage standardiserte kartleggingsoppgaver uten tidsbegrensning eller behov for at barna kan å lese eller skrive. I tillegg kan interaktive og spill-lignende elementer bidra til engasjement. I norsk kontekst kjenner vi til to nylig utviklede digitale verktøy for vurdering av barns matematiske kunnskap. Det ene verktøyet vurderer barnehagebarns tallkunnskap, geometri- og problemløsningsferdigheter gjennom motiverende og lekpregede oppgaver på nettbrett (ten Braak & Størksen, 2021). Det andre verktøyet, som vi har brukt i denne studien, består også av nettbrettoppgaver med spillelementer, men er spesifikt utviklet for å gi detaljert informasjon om barns tidlige tallforståelse – den typen tallforståelse som undervises i de to første skoleårene (Saksvik-Raanes & Solstad, 2021). For å komplementere vår eksisterende kunnskap om norske barnehagebarns tallforståelse, utnytter vi i denne studien muligheter i digitale verktøy til å danne et kvantitativt oversiktsbilde av barnehagebarns tallforståelse på slutten av barnehageårene og besvare følgende forskningsspørsmål:

Hvilket bilde av femåringers tallforståelse får vi fra vårt digitale kartleggingsverktøy for tidlig tallforståelse?

Bakgrunn

Tallforståelse

Det finnes flere ulike beskrivelser av hva som menes med begrepet tallforståelse (Berch, 2005; McIntosh et al., 1992). Beskrivelsene kan plasseres i tre ulike kategorier av tallforståelse (Whitacre et al., 2020): (i) en antatt medfødt mengdesans (*approximate number sense*), (ii) en tidlig tallforståelse (*early number sense*) spesifikt knyttet til de lærte ferdighetene og kunnskapene om tall som barnet tilegner seg i begynneropplæringen i barnehage og skole, og (iii) en viderekommen tallforståelse som blant annet handler om fleksibilitet innen ulike regneoperasjoner (*mature number sense*). I vår studie operasjonaliseres tallforståelse ved hjelp av rammeverket *Foundational Number Sense* (FoNS) (Andrews & Sayers, 2015) som hører til kategorien «tidlig tallforståelse» (Whitacre et al., 2020). FoNS-modellen beskriver ferdigheter knyttet til arbeid med tall i begynneropplæringen som barn ikke utvikler av seg selv (Andrews & Sayers, 2015).

FoNS-modellen er utviklet med utgangspunkt i en litteraturstudie som kategoriserte beskrivelser av barns utvikling av tallforståelse i det første året av sin formelle utdanning pluss/minus ett år. Litteraturstudien identifiserte åtte gjensidig avhengige tallforståelseskategorier: tallidentifikasjon, systematisk telling, tall og mengde, mengdediskriminering, forståelse for ulike representasjoner av tall, estimering, aritmetisk kompetanse og bevissthet om tallmønstre. Kategorien tallidentifikasjon handler om å identifisere tallsymbol og å kunne koble tallsymbol til verbale tallord. Systematisk telling innebærer å kunne telle systematisk og fleksibelt i intervallet 0 til 20, og inkluderer forståelse for ordinalitet. En bevissthet om forholdet mellom tallord og størrelsen på mengder innebærer å kunne koble tallord med en mengde som har samme kardinalitet som tallordet, og en forståelse for kardinalitet der det siste tallordet i en telling av en samling objekter angir antallet i mengden. Mengdediskriminering handler om en bevissthet om størrelsen av mengder, å kunne sammenligne mengder og å bruke begreper som større enn, mindre enn, flest og færrest. En forståelse for ulike representasjoner av tall innebærer en innsikt i at tall og mengder kan representeres på ulike måter ved hjelp av konkreter, fingre og tallinje. Estimering handler om å anslå størrelsen på en samling objekter eller plassering av et tall på en tallinje. Aritmetisk kompetanse innebærer å kunne transformere små mengder ved hjelp av addisjon eller subtraksjon. Bevissthet om tallmønstre medfører i FoNS-modellen å kunne gjenkjenne og utvide tallmønstre, eller avgjøre hvilket tall som mangler i en tallrekke (Andrews & Sayers, 2015).

De åtte kategoriene i FoNS antas å være gjensidig avhengig av hverandre og utgjør en sammensatt, fleksibel og relasjonell tallforståelse. Den gjensidige avhengigheten mellom kategoriene er et uttrykk for et helhetlig tallforståelsesperspektiv. Rammeverket er tidligere brukt for å analysere barns læringsmuligheter i ulike kulturelle kontekster (Andrews et al., 2015; Andrews & Sayers, 2015).

Tallforståelse i den norske begynneropplæringskonteksten

FoNS-modellen er utviklet med hensyn på en flerkulturell utdanningskontekst. Vi gir her et innblikk i den norske begynneropplæringskonteksten for deltakerne i vår studie, både i barnehagen og senere i skolen.

I Norge går 92,8 prosent av alle barn mellom ett og fem år i barnehage (Statistisk sentralbyrå, 2020). Det pedagogiske tilbudet i norske barnehager er styrt av *Rammeplan for barnehagen* (Kunnskapsdepartementet, 2017b). Rammeplanen sier at barnehagen skal legge til rette for barnets læring og utvikling, samtidig som det er barnets helhetlige utvikling og behov som skal ligge til grunn for aktivitetene i barnehagen. Gjennom sju kunnskapsområder beskriver rammeplanen læringsmuligheter barnehagen skal tilby gjennom sin pedagogiske virksomhet. Kunnskapsområdet *antall, rom og form* beskriver blant annet at barnehagen skal støtte barna i å danne erfaringer med å uttrykke tall, mengde og telling på ulike måter gjennom lek og eksperimenter (Kunnskapsdepartementet, 2017a). Disse kategoriene finner vi igjen som tallidentifikasjon, systematisk telling og tall og mengde i FoNS-modellen.

Opplæringstilbudet i den tiårige norske grunnskolen er styrt av Kunnskapsløftet (Kunnskapsdepartementet, 2019). Om vi ser spesifikt på kompetansemål relatert til tallforståelse etter 2. trinn, ser vi at elevene blant annet skal kunne telle, sammenligne og ordne tall på ulike måter, samt utforske tallbegreper med bruk av ulike representasjoner. Elevene skal også kunne plassere tall på tallinjen, bruke denne i regning og problemløsing, i tillegg til å utforske bruk av addisjon og subtraksjon gjennom praktisk erfaring (Kunnskapsdepartementet, 2019). Kompetansemålene kan plasseres i følgende kategorier i FoNS-modellen: tallidentifikasjon, systematisk telling, tall og mengde, mengdediskriminering og aritmetisk kompetanse.

Tidligere kunnskap om barnehagebarns tallforståelse

Fra et internasjonalt perspektiv vet vi en del om barns tallforståelse når de nærmer seg skolealder. Enkelte forskere har forsøkt å beskrive generelle utviklingsløp og hevder at de fleste amerikanske femåringer kan telle opp til 20–30 objekter, sammenligne mengder, identifisere tall i rekkefølge og bruke strategier til å finne løsningen på varierte problemtyper innen addisjon og subtraksjon med små mengder (Clements & Sarama, 2021). Tidligere studier har vist at amerikanske barn i barnehagealder kan løse problemløsningsoppgaver som inkluderer både addisjon, subtraksjon, divisjon og multiplikasjon ved hjelp av direkte modellering av situasjonene (Carpenter et al., 1993).

I norsk sammenheng finnes det få studier av barnehagebarns tallforståelse. Som en del av et longitudinelt prosjekt har Reikerås (2016) benyttet observasjonsmateriellet *MIO: matematikken, individet, omgivelsene* (Bergsmo et al., 2020; Davidsen et al., 2008), for å undersøke om barnehagebarn utvikler nødvendige matematiske kunnskaper før skolestart. Innenfor områder relatert til tallforståelse, viser studien at flertallet av fire- og femåringene i barnehagen mestret oppgavene relatert til telling og tallserier, mens et mindretall (5–15 %) fortsatt hadde utfordringer med å gjengi tallrekka opp til ti, og kardinalitetsprinsippet i telling til fem. Reikerås (2016) argumenterer for at barna viste et lavere nivå av tallforståelse enn jevnaldrende barn fra andre land. Samtidig er tallene i rimelig overensstemmelse med internasjonale anslag av andelen barn som opplever utfordringer i møte med matematikk (Geary, 2015).

Resultatene til Reikerås (2016) indikerer også at språk om mengder ikke er tilstrekkelig vektlagt i norske barnehager. En NOVA-rapport av Gulbrandsen og Eliassen (2013) støtter denne påstanden ved å vise at det er stor variasjon i hvor mye tid som vies til matematikk i ulike barnehager. Bare halvparten av de 649 barnehagene i studien rapporterte å ha «arbeidet ganske mye med» fagområdet *antall, rom og form*. Til sammenligning rapporterte åtti prosent det samme for fagområdet *kommunikasjon, språk og tekst*, noe som antyder at mange barnehager har et uforløst potensial i å integrere matematisk innhold i barnehagehverdagen.

Metode

Verktøy

Vi har utviklet et digitalt verktøy som skal hjelpe lærere med å beskrive tallforståelsen til fem og seks år gamle elever på starten av barneskolens 1. trinn i Norge (Saksvik-Raanes & Solstad, 2021). Tallforståelsesoppgavene presenteres i en webapplikasjon og er designet for å gjennomføres uten å lese eller skrive og uten direkte støtte fra en voksen. Applikasjonen gir taleforklaring til oppgavene og til hvordan applikasjonen fungerer. Oppgavesettet består av totalt 71 oppgaver fordelt over seks tallforståelseskategorier. Ni av oppgavene er dynamiske, som betyr at barnet skal flytte på objekter eller gjennomføre en prosess for å finne riktig svar, og 64 av oppgavene er statiske, som betyr at barnet skal svare ved å trykke på et svaralternativ. Appendiks 1 viser eksempler på et utvalg av oppgavene.

Hver oppgave ble skåret med ett poeng for rett svar og null poeng for feil svar. Verktøyet hadde ifølge Rasch-analysen en person reliabilitet på 0,87, som tilsvarer en Cronbach's alpha på 0,84 og betyr at oppgavene i stor grad måler samme konstrukt (tallforståelse).

Utvikling

I prosessen med å tilpasse oppgavenes design og vanskegrad til målgruppen ble verktøyet prøvd ut i en pilotstudie med barnehagebarn. Pilotstudien utgjør datagrunnlaget for analysen av barnehagebarns tallforståelse i denne artikkelen.

Verktøyet er utviklet med utgangspunkt i FoNS-modellen som er forankret i forskning på tidlig tallforståelse i begynneropplæringen i matematikk. I tillegg beskriver modellen tallforståelseskategorier som er mulige å operasjonalisere. Da studien ble gjennomført hadde vi utviklet oppgaver til de fem første kategoriene i FoNS-modellen: tallidentifikasjon, systematisk telling, tall og mengde, mengdediskriminering og aritmetisk kompetanse, i tillegg til kategorien subitisering. Disse kategoriene samsvarte i størst grad med beskrivelsene knyttet til tallforståelse i rammeplanen og Kunnskapsløftet. Underveis i datainnsamlingen ble oppgavesettet oppdatert fordi foreløpige analyser viste at flere av oppgavene var for lette. Til sammen er 71 ulike oppgaver inkludert i studien, men på grunn av oppdateringen ble hvert barn presentert for maksimalt 56 oppgaver.

Oppgavekategorier

Oppgavesettet inkluderte seks av FoNS-kategoriene som er tidligere beskrevet i bakgrunnskapittelet: tallidentifikasjon, systematisk telling, tall og mengde, mengdediskriminering, subitisering og aritmetisk kompetanse. Subitisering er lagt til som en kategori i våre undersøkelser ettersom ferdigheten er vektlagt i barns utvikling av FoNS (Sayers et al., 2016) og beskrevet som et område innen utvikling av tallforståelse som trenger mer oppmerksomhet (Clements et al., 2019). Å subitisere innebærer å raskt oppfatte og bestemme størrelsen av en mengde uten å telle. Man skiller mellom perseptuell subitisering, hvor man oppfatter og bestemmer størrelsen på en mengde uten å bruke andre matematiske prosesser, og konseptuell subitisering, hvor man visuelt deler opp mengdene i grupper og behandler gruppene med mental aritmetikk for å bestemme antallet (Clements et al., 2019). I subitiseringsoppgavene ble en mengde prikker vist i ett sekund før de forsvant og barna skulle angi hvor mange prikker de så.

Utvalg

Utvalget består av 77 barn fra fem barnehager i Trøndelag fylke; de fem første som meldte seg av 32 barnehager som mottok tilbud om å delta. De fem barnehagene var lokalisert i ulike bydeler og ulike kommuner i Trøndelag. Datainnsamlingen skjedde ved slutten av barnehageåret, omtrent to måneder før barna begynte på skolen. Deltakerne i studien nærmet seg dermed slutten på sin tid i barnehagen, og var mellom fem og seks år gamle på denne tiden. Vi refererer likevel til deltakerne som femåringer. Alle barn som var til stede i barnehagen på datainnsamlingsdagen fikk tilbud om og valgte å delta, uavhengig av bakgrunn og forutsetninger. Deltakerne representerer dermed et mangfold av barn med ulik språklig og sosial bakgrunn fra fem vilkårlig valgte trønderske barnehager, men er ikke et statistisk representativt utvalg av norske eller trønderske barnehagebarn.

Prosedyre

Med hjelp fra personalet organiserte vi grupper på to til fire barn som gjennomførte kartleggingen samtidig. Noen av barna gjennomførte også oppgavene individuelt. Barna fikk hvert sitt nettbrett og gjennomgikk samme instruks før det individuelle arbeidet med oppgavene startet. Instruksene omhandlet generelle sider ved opplegget, som at barna skulle prøve å svare på så mange oppgaver de ville på nettbrettet de fikk tildelt, og at de kunne avslutte arbeidet når de ønsket. Videre fikk barna mer spesifikke instrukser knyttet til funksjonalitetene i webapplikasjonen om hvordan de kom seg til neste oppgave og hvordan figuren i venstre hjørne på skjermen forklarte hva de skulle gjøre. De fleste barna gjorde alle oppgavene de ble presentert for, men noen hoppet over enkelte oppgaver eller valgte å avslutte før de hadde gjennomført alle oppgavene. Forskeren satt sammen med barna under hele datainnsamlingen og tok observasjonsnotater. Bakgrunn for observasjonsnotatene og videre analyse av dem er beskrevet under presentasjonen av det kvalitative datamaterialet.

Kvantitative analyser med Rasch målingsteori

For å beskrive generelle trekk ved barns tallforståelse ved skolestart er det hensiktsmessig å kunne måle tallforståelse. Når ikke alle barn har gjort alle de samme oppgavene er det problematisk å bruke totalt antall riktige svar eller andel riktige svar som mål på barnas tallforståelse. Dette problemet løses med å bruke Rasch målingsteori. I sosialvitenskapen kan vi kvantifisere egenskaper som i utgangspunktet er kvalitative ved hjelp av matematiske modeller (Stone, 1996). Rasch-analyser handler om å formelt teste data opp mot en statistisk modell for måling for å avgjøre hvor godt dataene passer med modellens forventninger til målingskonstruktet. I Rasch-modellen (Rasch, 1960) tilordnes hver person et mål på ferdighet (tallforståelse) og hver oppgave et mål på vanskegrad. Både ferdighet og vanskegrad måles på samme skala i enheten logit (Wright, 1977). Samme prinsipp brukes når vi måler ferdigheten til en høydehopper (hvor høyt hun kan hoppe) og vanskegraden til et høydestativ (hvor høyt stativet er) på samme skala i enheten meter.

Alle oppgavene i oppgavesettet skåres med rett eller galt, og vi har derfor brukt den dikotome Rasch-modellen for analysene. Modellen gir oss en sannsynlighet for at et barn med et gitt mål på tallforståelse vil svare riktig på en oppgave med en bestemt vanskegrad. Rasch-analysene beregner sannsynligheten for at barnet svarer rett på en oppgave ut fra forskjellen mellom barnets tallforståelse og oppgavens vanskegrad. I modellen har et barn med høyere tallforståelsesmål enn et annet barn større sannsynlighet for å svare riktig på en oppgave, uavhengig av hvilken oppgave i oppgavesettet det er snakk om. En oppgave med høyere vanskegrad enn en annen oppgave har lavere sannsynlighet for å bli besvart riktig, uavhengig av hvilket barn som løser oppgavene.

I en dikotom Rasch-modell, er X_{vi} en dikotom stokastisk variabel som vil si at X_{vi} har kun to mulige utfall, 0 eller 1. Sannsynligheten for at X_{vi} får utfallet 1, at person v svarer riktig på oppgave i, er gitt ved funksjonen

$$P\{X_{vi} = 1|\beta_v, \delta_i\} = \frac{e^{(\beta_v - \delta_i)}}{1 + e^{(\beta_v - \delta_i)}}$$

der, β_v angir personens ferdighet og δ_i angir oppgavens vanskegrad. Sannsynligheten for at person v får ett poeng på oppgave i er da avhengig av forskjellen mellom personens ferdighet, β_v, og oppgavens vanskegrad, δ_i. Sannsynligheten for at en person får ett poeng på en oppgave med samme vanskegrad som personens ferdighet er $P\ \{X_{vi} = 1 \,|\, \beta_v = \delta_i\} = 0{,}5$. Parametrene β_v og δ_i beregnes numerisk i en iterativ prosess i programvaren Winsteps (Linacre, 2017).

Vi har beskrevet tallforståelse med én variabel og dermed antatt at oppgavene måler samme endimensjonale konstrukt som i vårt tilfelle er tallforståelse. Dimensjonsanalyser av dataene underbygger at antakelsen om endimensjonalitet er meningsfull. Nesten alle oppgavene hadde en *outfit mnsq zstd* [1] under 1,75, noe som tilsier at de er produktive for måling med den endimensjonale Rasch-modellen vi har brukt (Wright & Linacre, 1994). For tre av oppgavene var *outfit mnsq zstd* over 2. Felles for disse oppgavene var at de var første oppgave i en ny oppgaveform, noe som indikerte at det kan være kvalitative sider ved barnas responsstruktur som var bakgrunnen for de høye *outfit*-verdiene.

En analyse av de standardiserte residualene viste at Rasch-dimensjonen forklarte 35,9 prosent av variansen i datamaterialet. Den uforklarte variansen i første kontrast var på 3,6 prosent med en eigenvalue på 3,9, noe som ikke nødvendigvis indikerer at tallforståelse er et flerdimensjonelt konstrukt. En *principal component analysis* (PCA) viste at enkelte av aritmetikk- og subitiseringsoppgavene grupperte seg i en egen kontrast, noe som kan indikere at disse danner en egen underdimensjon. Samtidig var ikke dette en entydig gruppering da noen av oppgavene innen aritmetikk og subitisering fordelte seg på en annen måte. Forhold som oppgavenes vanskegrad og utforming kan ha påvirket barnas svar. Ettersom våre analyser ikke viser klare tegn på at enkelte tallforståelseskategorier skiller seg ut ser vi på tallforståelse som én latent variabel.

Kvalitativt datamateriale

Det kvalitative datamaterialet består av notater fra enkeltobservasjoner gjort mens barna arbeidet med tallforståelsesoppgavene. Hensikten med observasjonene var å se om enkelte barn brukte løsningsstrategier vi ikke hadde forutsett og som kunne ha betydning for tolkningen av de kvantitative resultatene. Når noen av barna kom med ytringer om hvordan de forstod eller løste oppgavene med en klart uttalt strategi, ble dette samlet i observasjonsnotatene. Etter datainnsamlingen i hver barnehage ble det skrevet en mer fullstendig rapport hvor ulike forhold i datainnsamlingen knyttet til barnas interaksjon med oppgavene ble oppsummert. Rapportene og observasjonsnotatene ble analysert for å identifisere bruk av strategier og utsagn som kunne si noe om barnas tidligere erfaringer med oppgavekonteksten. For å belyse sider ved kartleggingsprosessen som de kvantitative dataene ikke kunne fange opp, har vi inkludert kvalitative beskrivelser av ett barns løsningsstrategi og sammenstilt dem med barnets kvantitative resultater på oppgavesettet. Dette er ikke en beskrivelse av et generelt aspekt ved gruppen, men en illustrasjon og et eksistensbevis på elementer som bør tas i betraktning når man tolker kvantitative data for enkeltpersoner.

1 Outfit MNSQ Zstd er forenklet sett et mål på hvor mye svarene på en oppgave avviker fra Rasch-modellens forventede svar, uttrykt i antall standardavvik fra forventningsverdien (Zstd). Verdier langt unna null betyr at mange barn ikke har svart det man forventet på oppgaven ut fra deres tallforståelsesmål. Slike oppgaver kan ha elementer som ikke er relatert til tallforståelse men forstyrrer eller hjelper barnet. Outfit er et mål som vekter overraskende svar mest og mean squares (MNSQ) betyr at man kvadrerer avvikene før gjennomsnittet beregnes slik at store avvik vektes mer.

Etiske betraktninger

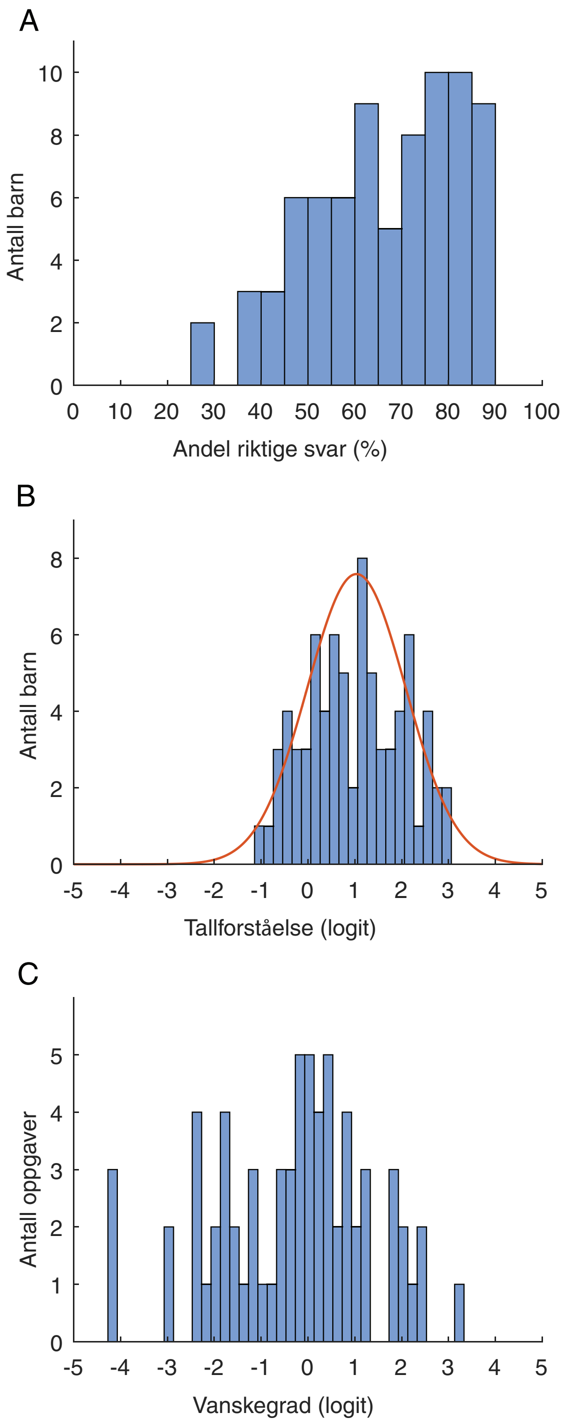
I møte med barna i barnehagene var det sentralt å gjennomføre datainnsamlingen på barnas premisser, i tråd med retningslinjer utgitt av Den nasjonale forskningsetiske komité for samfunnsvitenskap og humaniora (2021). Barna fikk sitte på et kjent rom i barnehagen, og en av de voksne i barnehagen var tilgjengelig for barna under arbeidet. Om barn a var usikre på svaret på enkeltoppgaver underveis ble de oppfordret av forskeren til å trykke på det svaret de trodde var riktig. Barna fikk avslutte arbeidet med oppgavene når de selv ønsket det. De fleste barna valgte å fullføre aktiviteten, og alle fikk hjelp til å komme inn i barnehagehverdagen igjen når de var ferdige med oppgavene. Barnas oppgavesvar ble innsamlet anonymt på nettbrett som forskeren hadde med seg. Oppgavesvarene ble registrert som rett eller feil gjennom en digital plattform uten videre innsamling av personidentifiserende informasjon.

I dialog med Norsk senter for forskningsdata (NSD) ble det før datainnsamlingen avklart at studien ikke utløste meldeplikt hos NSD, og at det ikke var nødvendig med samtykke ut over å informere foresatte om studien da studien ikke samler inn personidentifiserende opplysninger. Informasjon til foresatte ble sendt ut via barnehagene.

Resultater

Mange barn går ut fra barnehagen med godt utviklet tallforståelse

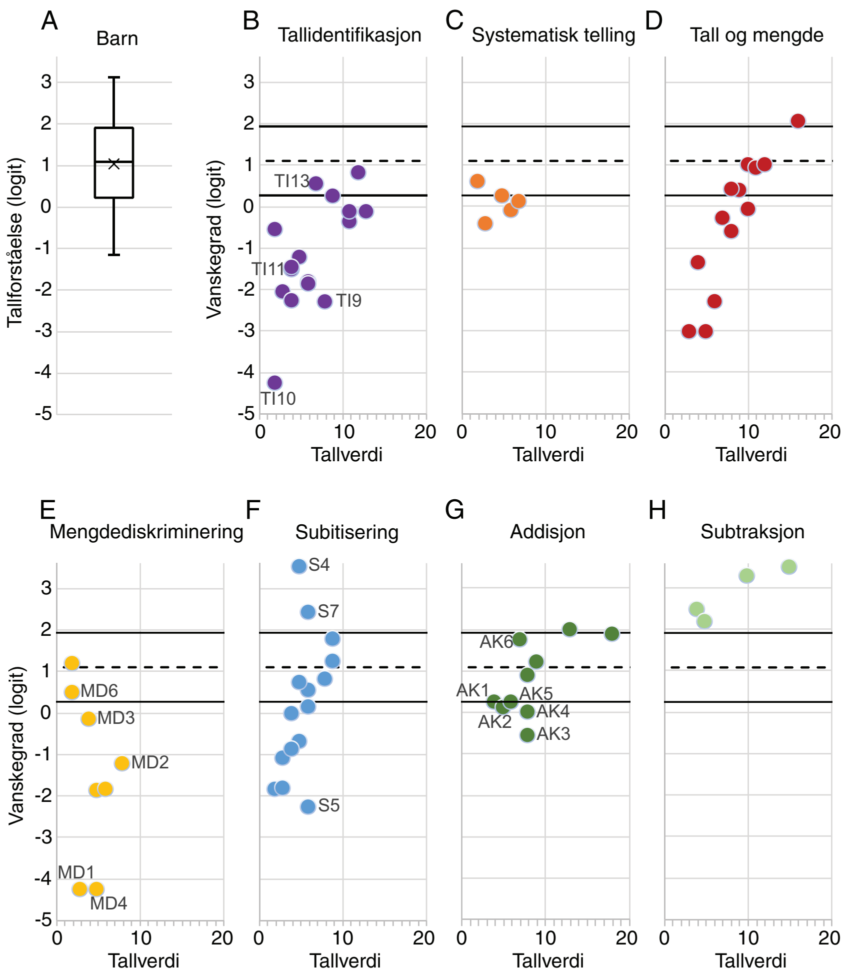
Som gruppe viste barnehagebarna i denne studien overraskende godt utviklet tallforståelse, spesielt ettersom oppgavesettet var laget ut fra beskrivelsene i FoNS-rammeverket som er tilpasset matematikken barn møter i 2. trinn i Norge. Figur 1A viser at andelen riktige svar på oppgavene var skjevfordelt mot høye andeler riktige svar. Alle femåringene fikk riktig svar på minst 25 prosent av oppgavene, og hovedvekten av barna fikk til rundt 80 prosent av oppgavene. Denne skjevfordelingen betyr imidlertid ikke at tallforståelse var skjevfordelt i utvalget, men oppstår fordi det var mange flere enkle oppgaver enn vanskelige oppgaver. For å få et riktig inntrykk av fordelingen av barnas tallforståelse brukte vi Rasch-analyse som tar hensyn til at oppgavene har ulik vanskegrad. Da så vi at barnas tallforståelse var tilnærmet normalfordelt (figur 1B; Kolmogorov-Smirnov test for normalitet; $p = 0{,}2$; $df = 79$). Rasch-analysen gir oss også vanskegraden til oppgavene, som er på samme måleskala som, og kan sammenlignes direkte med, barnas tallforståelse i figur 1C. Vi ser at tjue oppgaver hadde lavere vanskegrad enn alle barnas tallforståelsesmål. Dette var oppgaver som handlet om å identifisere og koble tallsymboler og mengder på inntil fem objekter. Videre lå barnas gjennomsnittlige tallforståelse (1 logit) over vanskegraden til de aller fleste oppgavene (57 av 71), og en fjerdedel av barna hadde et tallforståelsesmål over 1,9 logit, som er vanskegraden til addisjonsoppgaver med tosifrede tallsymboler.



Figur 1. Tallforståelse og vanskegrad. A. Fordeling av barnas andel riktige svar; B. Fordeling av barnas tallforståelse sammenlignet med normalfordelingen (rød kurve); C. Fordeling av oppgavenes vanskegrad

FoNS-kategori

For å få et nærmere innblikk i hva som kjennetegnet tallforståelsen til barna i denne studien har vi sammenlignet barnas tallforståelse med vanskegraden til hver enkelt oppgave i hver FoNS-kategori i figur 2. Barnas tallforståelse er delt inn i kvartiler og vist som horisontale linjer i hvert panel for å forenkle sammenligningen med oppgavenes vanskegrad. Hvert panel i figur 2 viser vanskegraden til oppgavene i en FoNS-kategori, sortert etter største tallverdi i oppgaven.



Figur 2. Mål på barnas tallforståelse og oppgavenes vanskegrad i enheten logit. A. Boksplott av barnas tallforståelse (y-aksen). Gjennomsnittet er markert med x. B-H. Spredningsplott av oppgavenes vanskegrad (y-aksen) for hver FoNS-kategori. Oppgavene er ordnet etter største tallverdi i oppgaven (x-aksen). Hel linje markerer første og tredje kvartil for barnas tallforståelsesmål. Stripete linje markerer median for barnas tallforståelsesmål. Oppgaver som er nevnt i teksten er markert med navn i hvert enkelt spredningsplott (TI9, TI10, TI11 og TI13 i kategorien Tallidentifikasjon).

I det følgende ser vi nærmere på tre områder som påvirket vanskegraden til oppgavene og dermed forteller noe om barnas tallforståelse: FoNS-kategori, tallområde og matematisk språk.

Tallidentifikasjon, tall og mengde og mengdediskriminering var kategorier som hadde lav vanskegrad i forhold til barnas tallforståelse. Aritmetisk kompetanse var som ventet den vanskeligste FoNS-kategorien. Subtraksjon skilte seg ut som spesielt vanskelige oppgaver, men seks barn (8 %) svarte riktig på de vanskeligste av disse oppgavene. Alle de letteste aritmetikkoppgavene (AK1–AK5, se eksempler i appendiks 1) involverte objekter arrangert i mønstre som vi antok var lett gjenkjennbare for mange av barna, som for eksempel terningmønster. AK6 var representert bare med verbal instruks, mens de vanskeligste oppgavene var representert med symboler. Det kan altså se ut til at overgangen fra konkrete mengder til symboler utgjorde et skille i tallforståelse hos barna med høyest tallforståelse. Omtrent 75 prosent av barna svarte riktig på oppgaver om å addere konkrete mengder på til sammen inntil ti objekter. Omtrent 25 prosent av barna svarte riktig på addisjon med symboler og tallverdi over ti.

Subitiseringsoppgaver med mengder større enn fem var omtrent like vanskelige som addisjonsoppgaver med symboler (figur 2 F–H). Dette kan bety at barna bruker mental addisjon for å løse slike konseptuelle subitiseringsoppgaver som krever at barnet utnytter at en mengde kan deles inn i delmengder (Clements et al., 2019; Starkey & McCandliss, 2014).

Tallområde

Alle barnehagebarna i denne studien kunne løse oppgaver med tall og mengder opp til og med fem i fire av FoNS-kategoriene (figur 2). Dette komplementerer studien til Reikerås (2016), hvor omtrent femten prosent av barna ikke ble observert i å dekke på til fem personer eller gjengi hvor mange objekter det er i en mengde etter å ha telt fem objekter («kardinalitetsprinsippet»). Oppgavesettet har ikke mange oppgaver med tallverdier over ti, men resultatene viser at over halvparten av barna hadde forståelse for tall over ti.

I noen, men ikke alle, av FoNS-kategoriene ble oppgavene vanskeligere dess større tallverdiene i oppgaven var. For tallidentifikasjon var korrelasjonen mellom tallverdi og vanskegrad $r = 0{,}7$ ($p < 0{,}01$), men hvis vi bare ser på oppgavene med identifikasjon av formelle tallsymboler ser det ut til at vanskegraden i hovedsak var bestemt av om tallsymbolene var større eller mindre enn ti.

For tall og mengde var korrelasjonen mellom tallverdi og vanskegrad $r = 0{,}89$ ($p < 0{,}001$). For subitisering var korrelasjonen ikke signifikant, men tre av oppgavene skilte seg ut ved å være spesielt enkle (S5: terningmønster for 6) eller spesielt vanskelige (S7 og S4: en kolonne med fem prikker som det er kjent at er vanskelig å oppfatte). Uten disse tre oppgavene var korrelasjonen mellom tallverdi og vanskegrad $r = 0{,}91$ ($p < 0{,}001$).

Matematisk språk og begreper

Resultatene fra kategorien tallidentifikasjon viste at de fleste barnehagebarna var kjent med formelle tallsymboler, men ikke uformelle symboler som tellestreker. Mens tellestreker for tallene 2 (TI10) og 4 (TI11) var oppgaver alle barna fikk til, var det å gjenkjenne tellestreksymbolet for mengden 7 (TI13; 0,52 logit) nesten 3 logit vanskeligere enn å gjenkjenne det formelle tallsymbolet for 8 (TI9; -2,84 logit) og 1 logit vanskeligere enn å knytte tallsymbolet 7 til den tilsvarende mengden (TM2; -0,5 logit). Resultatet kan tolkes dit hen at omtrent en tredjedel av barna ikke kjente betydningen av den horisontale tellestreken i symbolet for 5, og antyder at disse barna hadde begrenset erfaring med opptellings-aktiviteter og spill hvor denne strategien brukes for å holde orden på antall over 5.

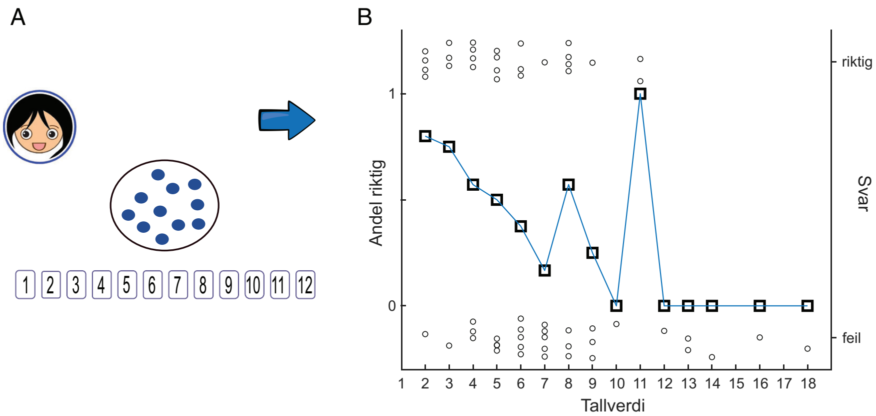
Terninger er en annen form for uformelle tallsymboler som barn møter i sammenheng med spill. Kategorien subitisering hadde oppgaver der noen mengder var strukturert som terningmønster. Alle disse terningmønstrene hadde vanskegrad på -0,7 logit eller lavere, som er lavere enn 76 av barna sine tallforståelsesmål. Til sammenligning hadde oppgaver med tilsvarende mengder, men som ikke var strukturert som et ikonisk mønster, vanskegrader høyere enn -0,11 logit. At innlærte, ikoniske mønstre er lettere å subitisere er kjent fra litteraturen (Leibovich-Raveh et al., 2018). Selv om forskjellen mellom terningmønstre og ustrukturerte mønstre ikke er systematisk undersøkt i denne studien, tyder resultatene på at mange av barna i studien hadde kjennskap til ikoniske mønstre som er forbundet med terningspill.

Oppgavene innen systematisk telling skiller seg ut i figur 2 med lite spredning og høy vanskegrad i forhold til tallverdi, og var heller ikke korrelert med tallverdi. En mulig årsak til dette er at alle oppgavene i denne kategorien inneholder ord knyttet til ordinalitet, som «før», «etter», «andre», «tredje» og «femte». Bare omtrent halvparten av barna i denne studien viste forståelse for disse begrepene og ordene. Til sammenligning viste omtrent nitti prosent av femåringene i studien til Reikerås (2016) at de kunne identifisere objekter «i midten» av en rekke objekter, som også handler om språk for ordinalitet.

Et tilsvarende fenomen ser vi i kategorien antallsdiskriminering. Oppgavene som inneholder ordene «færrest» (0,28 logit; MD3) og «færre enn» (0,26 logit; MD6) hadde betydelig høyere vanskegrad enn oppgavene som inneholder ordene «flest» (-1,25 logit; MD2) og «flere enn» (-2,61 logit; MD4) til tross for at de to sistnevnte oppgavene hadde høyere tallverdi. Resultatene kan tyde på at barna i studien ikke hadde like mye erfaring med begrepene «færrest» og «færre enn». Dataene i denne studien ble samlet inn i en landsdel hvor man i dagligtale bruker ordet «mindre» om både antall og mål og ordet «færre» ikke er en del av dialekten (trøndersk). Det er altså mulig at flere barn hadde forståelse for *begrepene* «færre enn» og «færrest» enn det kan se ut som i denne studien selv om de ikke kjente ordene.

Kvalitative observasjoner knyttet til barnas tallforståelse

De kvantitative resultatene vi har presentert hittil viser hva deltakerne i studien har svart på de ulike oppgavene. Samtidig er det sider ved barnas tallforståelse, som barnas strategier og



Figur 3. Petters løsning. A. Oppgaven TM8 fra FoNS-kategorien Tall og mengde. Barna fikk den verbale instruksen «Trykk på tallet som viser hvor mange baller det er i rundingen». B. Andel riktige svar på oppgavene som funksjon av største tall i oppgaven for Petter som resonnerte seg frem til riktig tallsymbol for 11. Svarte kvadrater angir andel riktige svar (venstre y-akse) for hver tallverdi (x-aksen), mens svarte sirkler angir feil eller riktig svar for hver oppgave (høyre y-akse).

løsningsprosess, som de kvantitative dataene ikke kan fange opp. For å illustrere noe av den matematiske kunnskapen som de digitale oppgavene *ikke* fanger opp, gir vi her en kvalitativ beskrivelse av Petters løsning av oppgave TM8 (figur 3A). Episoden viser at selv om et barn ikke har svart riktig på oppgaver med tall over ti, kan barnet likevel ha god forståelse for telling og være i stand til å resonnere om tall som er utenfor sitt eget tallområde.

Petter målte 0,15 logit på tallforståelseskalaen. Figur 3B viser at Petter fikk få riktige svar på oppgaver som involverte tall større enn fem. Samtidig ser vi at oppgaver med tallene 8 og 11 som svar skiller seg ut som oppgaver han mestrer i større grad. Observasjoner av Petter da han løste oppgave TM8, viste at han benyttet seg av peketelling for å bestemme antall objekter i sirkelen og endte på 11 i sin verbale tallrekke. Petter visste altså at det siste tallet i tellesekvensen ga han kardinaliteten til mengden, men han visste ikke hvordan tallsymbolet 11 så ut. Etter å ha tenkt seg litt om, kom Petter på at han kunne gjenkjenne 7-tallet. Dermed fant han tallet 7 blant rekka med tallsymboler, som var oppgitt som svaralternativer, og telte seg oppover fra 7 for å finne symbolet for 11.

Diskusjon

Femåringenes tallforståelse sammenlignet med tidligere kunnskap

Kunnskap om barns tallforståelse kan legge føringer for hva slags matematikkaktiviteter barnehagene legger til rette for. Tidligere studier har antydet at enkelte norske barnehagebarns tallforståelse utvikles senere enn det vi ser internasjonalt (Reikerås, 2016). Vi mangler imidlertid kunnskap om tallforståelsen til bredden av norske barnehagebarn og spesielt til de som kan mest om tall. Til nå har det også vært ressurskrevende å samle denne typen kunnskap. I denne studien har vi sett at digitale oppgaver kan gi oss et oversiktsbilde av tallforståelsen til femåringer i norske barnehager på en mer tidseffektiv måte enn tidligere.

Vi fant at det var betydelig spredning i barnas tallforståelse, fra barn som viste begrenset forståelse for tallsymboler og mengder over fem, til barn som behersket aritmetikk med tosifrede tallsymboler. Som gruppe viste barna vesentlig større grad av tallforståelse enn det man kunne forvente ut fra tidligere forskning og rammeplanen for barnehagens beskrivelser av tallidentifikasjon, systematisk telling og forhold mellom tall og mengde. Eksempelet med Petters løsning viser også at barn som ikke har lært alle tallsymbolene, likevel kan ha tilstrekkelig forståelse for tallsystemet til å resonnere seg frem til riktig tallsymbol ved å utnytte strukturen i kjente representasjoner. I sum kan vi nå si at mens enkelte femåringer har behov for praktiske erfaringer med tallmengder inntil fem, har mange barn allerede bred forståelse for mengder og tall til og ut over ti. For å engasjere denne potensielt store gruppen barn i matematiske spørsmål og tenkemåter, vil det være nødvendig å tilby dem utfordringer i møte med matematikken både i barnehagen og senere i skolen (Clements et al., 2013; Rinvold, 2017).

Barnehagen kan være en viktig arena for barnets utvikling av tidlig tallforståelse. En nylig publisert intervensjonsstudie antyder at flere barnehager allerede har en systematisk tilnærming til matematikk, og at strukturerte læringsaktiviteter kombinert med kompetanseheving hos personalet kan være positivt for barns matematiske utvikling (Rege et al., 2021). Samtidig er det stor variasjon mellom barnehagene med tanke på hvilke aktiviteter det legges til rette for (Gulbrandsen & Eliassen, 2013; Urban et al., 2022). Våre data gir ikke grunnlag for å si noe om hvor barna har gjort sine matematiske erfaringer eller hvorfor det er stor spredning i barnas tallforståelse. Det vil derfor være av interesse å se nærmere på barnehagens rolle i å tilby barn relevante matematiske erfaringer og utjevne forskjeller i tråd med rammeplanen (Kunnskapsdepartementet, 2017b) i videre forskning.

Norsk forskning i barnehagefeltet har med et prosessorientert fokus satt søkelyset på samhandling mellom barn og voksne i barnehagen (Breive, 2020; Nergård, 2021) og kompleksiteten i å lede matematiske diskusjoner (Justnes & Mosvold, 2021) der barnehagelæreren må improvisere og løse flere pedagogiske og matematiske utfordringer for å videreutvikle barnets lek og spontane hverdagssituasjoner til læring (Sæbbe & Mosvold, 2020). Barnehagelærerens kompetanse kan dermed være av stor betydning for den matematiske samhandlingen i barnehagen (Hundeland et al., 2020). Kunnskap om variasjonen i barnehagebarns tallforståelse kan bidra til at barnehagelærere vier mer oppmerksomhet til viktigheten av å tilpasse innholdet i matematisk lek og samhandling til barnas kunnskaper. På denne måten kan våre kvantitative resultater indirekte bidra til å heve kvaliteten på de matematiske prosessene som foregår i en kompleks barnehagehverdag.

Muligheter i det digitale formatet og videre forskning

Vårt øyeblikksbilde av femåringers tallforståelse gjennom digitale oppgaver utfyller de longitudinelle observasjonsstudiene til Reikerås (2016). Studiene tegner et sammensatt bilde av norske barnehagebarns tallforståelse. Videre vil det kreve et samspill mellom kvalitative observasjoner av barns tallforståelse i hverdagen og kvantitative øyeblikksbilder for å få mer utfyllende kunnskap.

Det digitale formatet muliggjør undersøkelser av en større gruppe barns tallforståelse på en måte som tilfredsstiller kravene til validitet (Saksvik-Raanes & Solstad, 2021). Samtidig viser våre kvalitative resultater at det er sider ved barns tallforståelse, som barnas strategier og løsningsprosesser, som de kvantitative dataene ikke har fanget opp. Det er fortsatt et stort uforløst potensial i digitale kartlegginger med spill-elementer som kan bidra til å styrke ulike aspekter ved kartleggingenes validitet. For eksempel hadde enkelte av oppgavene i oppgavesettet vi benyttet en større grad av interaktivitet, der barna skulle strukturere objekter på skjermen for å finne en løsning. Barnas løsningsprosess kan spilles av og analyseres i etterkant. Denne muligheten åpner for å få dypere innsikt i barnas løsningsstrategier, men hvordan denne informasjonen best kan tolkes og benyttes må avklares i egne studier. Der vår studie var begrenset til et øyeblikksbilde av barnas tallforståelse, åpner inntoget av digitale matematikkaktiviteter som integrerer lek og læring i hjem og barnehage en ny kilde til kunnskap om hvordan barn utvikler og anvender sin tallforståelse (Carlsen et al., 2016; ten Braak & Størksen, 2021).

Med bakgrunn i rammeplanens beskrivelser av barnets læringsmuligheter, helhetlige utvikling og behov, er det forbundet en skepsis til kartlegging av barnehagebarns kompetanse ut fra et humanistisk syn på læring og utvikling (Vik, 2017). Den digitale utviklingen åpner opp for å samle mer informasjon om barns forståelse gjennom lystbetonte aktiviteter på nettbrett, noe som kan føre til utvikling av ny kunnskap om barns tidlige matematiske kompetanse. I denne sammenhengen er det viktig å være påpasselig med at resultatene fra kartlegginger blir brukt i samsvar med det nordiske verdigrunnlaget i barnehagen (Urban et al., 2022). Samtidig er kunnskap om barns matematiske kompetanse et nødvendig grunnlag for å kunne gi alle barn like muligheter til å lære, bruke og glede seg over matematikk, både i barnehagen og videre i livet.

Författarbiografi

Gunnhild Saksvik-Raanes er doktorgradsstipendiat i matematikkdidaktikk ved Institutt for lærerutdanning, NTNU. Temaet for doktorgraden er digital vurdering av barns tallforståelse i overgangen fra barnehage til skole. Gunnhild har tidligere jobbet som lærer i grunnskolen med særlig interesse for begynneropplæringen i matematikk.

Trygve Solstad er førsteamanuensis i matematikkdidaktikk ved Institutt for lærerutdanning, NTNU, Norge. Han har sin ph.d. i nevrovitenskap fra NTNU og har forsket på matematiske modeller for læring og hukommelse som postdoktor ved the Salk Institute, USA.

Han er nå leder for forskergruppen Matematisk kognisjon og literacy ved NTNU som forsker på tallforståelse og matematiske læringsprosesser.

Yvonne Grimeland er førsteamanuensis i matematikkdidaktikk ved Institutt for lærerutdanning, NTNU. Hun har en ph.d. i matematikk, innen abstrakt algebra, fra NTNU. Hun har tidligere undervist i matematikkdidaktikk ved Dronning Mauds Minne høgskole for barnehagelærerutdanning, og er særskilt interessert i yngre barns arbeid med matematikk.

Referanser

Andrews, P. & Sayers, J. (2015). Identifying opportunities for grade one children to acquire foundational number sense: Developing a framework for cross cultural classroom analyses. *Early Childhood Education Journal*, *43*(4), 257–267. https://doi.org/10.1007/s10643-014-0653-6

Andrews, P., Sayers, J. & Marschall, G. (2015). Developing foundational number sense: Number line examples from Poland and Russia. I K. Krainer & N. Vondrová (Red.), *Proceedings of the Ninth Congress of the European Society for Research in Mathematics Education* (s. 1681–1687). European Society for Research in Mathematics Education.

Benoit, L., Lehalle, H., Molina, M., Tijus, C. & Jouen, F. (2013). Young children's mapping between arrays, number words, and digits. *Cognition, 129*(1), 95–101. https://doi.org/10.1016/j.cognition.2013.06.005

Berch, D. B. (2005). Making sense of number sense: Implications for children with mathematical disabilities. *Journal of Learning Disabilities, 38*(4), 333–339. https://doi.org/10.1177/00222194050380040901

Bergsmo, H. S., Dalvang, T. & Kirsti, L. R. E. (2020). *MIO. Matematikken. Individet. Omgivelsene. Håndbok* (2. utg.). GAN Aschehoug.

Breive, S. (2020). Student–teacher dialectic in the co-creation of a zone of proximal development: An example from kindergarten mathematics. *European Early Childhood Education Research Journal*, *28*(3), 413–423. https://doi.org/10.1080/1350293X.2020.1755498

Carlsen, M., Erfjord, I., Hundeland, P. S. & Monaghan, J. (2016). Kindergarten teachers' orchestration of mathematical activities afforded by technology: Agency and mediation. *Educational Studies in Mathematics, 93*(1), 1–17. https://doi.org/10.1007/s10649-016-9692-9

Carpenter, T. P., Ansell, E., Franke, M. L., Fennema, E. & Weisbeck, L. (1993). Models of problem solving: A study of kindergarten children's problem-solving processes. *Journal for Research in Mathematics Education JRME, 24*(5), 428–441. https://doi.org/10.5951/jresematheduc.24.5.0428

Clarke, B., Cheeseman, J. & Clarke, D. (2006). The mathematical knowledge and understanding young children bring to school. *Mathematics Education Research Journal, 18*(1), 78–102. https://doi.org/10.1007/BF03217430

Clarke, D., Clarke, B. & Roche, A. (2011). Building teachers' expertise in understanding, assessing and developing children's mathematical thinking: The power of task-based, one-to-one assessment interviews. *ZDM, 43*(6), 901–913. https://doi.org/10.1007/s11858-011-0345-2

Clements, D. H., Baroody, A. J. & Sarama, J. (2013). *Background research on early mathematics*. National Governor's Association.

Clements, D. H., Sarama, J. & MacDonald, B. L. (2019). Subitizing: The neglected quantifier. I A. Norton & M. Alibali (Red.), *Constructing number. Research in Mathematics Education* (s. 13–45). Springer. https://doi.org/10.1007/978-3-030-00491-0_2

Clements, D. H., Sarama, J., Tatsuoka, C., Banse, H. & Tatsuoka, K. (2021). Evaluating a model for developing cognitively diagnostic adaptive assessments: The case of young children's length measurement. *Journal of Research in Childhood Education*, 1–16. https://doi.org/10.1080/02568543.2021.1895921

Clements, D. & Sarama, J. (2021). *Learning and teaching early math: The learning trajectories approach* (3. utg.). Routledge.

Davidsen, H. S., Løge, I. K., Lunde, O., Reikerås, E. & Dalvang, T. (2008). *MIO. Matematikken. Individet. Omgivelsene. Håndbok.* Aschehoug.

Den nasjonale forskningsetiske komité for samfunnsvitenskap og humaniora. (2021). *Forskningsetiske retningslinjer for samfunnsvitenskap og humaniora*. https://www.forskningsetikk.no/

Engel, M., Claessens, A. & Finch, M. A. (2013). Teaching students what they already know? The (mis) alignment between mathematics instructional content and student knowledge in kindergarten. *Educational Evaluation and Policy Analysis, 35*(2), 157–178. https://doi.org/10.3102/0162373712461850

Geary, D. C. (2015). The classification and cognitive characteristics of mathematical disabilities in children. I R. C. Kadosh & A. Dowker (Red.), *The Oxford handbook of numerical cognition* (s. 767–786). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199642342.013.017

Ginsburg, H. P. (2002, 21.–26. juli). Little children, big mathematics: Learning and teaching in the preschool. I *Proceedings of the 26th Conference of the International Group for the Psychology of Mathematics Education*. https://files.eric.ed.gov/fulltext/ED476078.pdf

Ginsburg, H. P. & Pappas, S. (2016). Invitation to the birthday party: Rationale and description. *ZDM, 48*, 947–960. https://doi.org/10.1007/s11858-016-0818-4

Gulbrandsen, L. & Eliassen, E. (2013). *Kvalitet i barnehager. Rapport fra en undersøkelse av strukturell kvalitet høsten* 2012 (NOVA-rapport 1/13). https://hdl.handle.net/20.500.12199/5062

Hogsnes, H. D. & Moss, T. (2014). Forståelser av gode overganger og opplevelse av sammenheng mellom barnehage, skole og skolefritidsordning. *Nordic Early Childhood Education Research Journal, 7*(6), 1–24. https://doi.org/10.7577/nbf.625

Hundeland, P. S., Carlsen, M. & Erfjord, I. (2020). Qualities of mathematical discourses in kindergartens. *ZDM, 52*(4), 691–702. https://doi.org/10.1007/s11858-020-01146-w

Justnes, C. N. & Mosvold, R. (2021). The work of leading mathematical discussions in kindergarten: A Norwegian case study. I G. A. Nortvedt, N. Buchholtz, J. Fauskanger, F. Hreinsdóttir, M. Hähkiöniemi, B. E. Jessen, J. Kurvits, Y. Liljekvist, M. Misfeldt, M. Naalsund, H. K. Nilsen, G. Pálsdóttir, P. Portaankorva-Koivisto, J. Radišić & A. Wernberg (Red.), *Bringing Nordic mathematics education into the future. Proceedings of Norma 20. The ninth Nordic Conference on Mathematics Education* (s. 153–160). Swedish Society for Research in Mathematics Education.

Kunnskapsdepartementet. (2017a). *Antall, rom og form.* https://www.udir.no/laring-og-trivsel/rammeplan-for-barnehagen/fagomrader/antall-rom-form/

Kunnskapsdepartementet. (2017b). *Rammeplan for barnehagen. Forskrift om rammeplan for barnehagens innhold og oppgaver*. https://www.udir.no/laring-og-trivsel/rammeplan-for-barnehagen/

Kunnskapsdepartementet. (2019). *Læreplan i matematikk 1.–10. trinn* (MAT01-05). https://www.udir.no/lk20/mat01-05?lang=nob

Leibovich-Raveh, T., Lewis, D. J., Al-Rubaiey Kadhim, S. & Ansari, D. (2018). A new method for calculating individual subitizing ranges. *Journal of Numerical Cognition, 4*(2), 429–447. https://doi.org/10.5964/jnc.v4i2.74

Lembrér, D. & Meaney, T. (2016). Preschool children learning mathematical thinking on interactive tables. I T. Meaney, O. Helenius, M. L. Johansson, T. Lange & A. Wernberg (Red.), *Mathematics Education in the Early Years* (s. 235–254). Springer. https://doi.org/10.1007/978-3-319-23935-4_13

Linacre, J. M. (2017). *Winsteps® Rasch measurement computer program. User's guide*. Winsteps.

Lopez-Pedersen, A., Mononen, R., Korhonen, J., Aunio, P. & Melby-Lervåg, M. (2021). Validation of an early numeracy screener for first graders. *Scandinavian Journal of Educational Research, 65*(3), 404–424. https://doi.org/10.1080/00313831.2019.1705901

McIntosh, A., Reys, B. J. & Reys, R. E. (1992). A proposed framework for examining basic number sense. *For the Learning of Mathematics, 12*(3), 2–8,44.

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

Nergård, B. (2021). Preschool children's mathematical arguments in play-based activities. *Mathematics Education Research Journal*. https://doi.org/10.1007/s13394-021-00395-6

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. University of Chicago Press.

Rege, M., Størksen, I., Solli, I., Kalil, A., Mcclelland, M., ten Braak, D., Lenes, R., Lunde, S., Breive, S., Carlsen, M., Erfjord, I. & Hundeland, P. (2021). The effects of a structured curriculum on preschool effectiveness: A field experiment. *Journal of Human Resources*, 0220-10749R3. https://doi.org/10.3368/jhr.0220-10749R3

Reikerås, E. (2016). Central skills in toddlers' and pre-schoolers' mathematical development, observed in play and everyday activities. *NOMAD. Nordic Studies in Mathematics Education, 21*(4), 57–78.

Rinvold, R. A. (2017). To streker under svaret – matematiske møter mellom teori og praksis. I M. Løtveit (Red.), *Tidssignaler: Lærerutdanningsfag i utvikling: Utdanning av lærere på Hamar – 150 år* (s. 259–276). Oplandske Bokforlag.

Rinvold, R. A. & Erstad, H. (2015). Learning interventions supporting numerosity in three year old children. I K. Kraine & N. Vondrová (Red.), *Proceedings of the ninth congress of the European Society for Research in Mathematics Education* (s. 1968–1974). ERME. https://hal.archives-ouvertes.fr/hal-01288494

Saksvik-Raanes, G. & Solstad, T. (2021). Developing a formative, teacher-oriented, digital tool to assess number sense in school starters. I G. A. Nortvedt, N. F. Buchholtz, J. Fauskanger, M. Hähkiöniemi, B. E. Jessen, M. Naalsund, H. K. Nilsen, G. Pálsdóttir, P. Portaankorva-Koivisto, J. Radišić, J. Ö. Sigurjónsson, O. Viirman & A. Wernberg (Red.), *Bringing Nordic mathematics education into the future. Proceedings of Norma 20. The ninth Nordic conference on mathematics education* (s. 212–219). Swedish Society for Research in Mathematics Education.

Sayers, J., Andrews, P. & Boistrup, L. B. (2016). The role of conceptual subitising in the development of foundational number sense. I T. Meaney, O. Helenius, M. L. Johansson, T. Lange & A. Wernberg (Red.), *Mathematics education in the early years* (s. 371–394). Springer. https://doi.org/10.1007/978-3-319-23935-4_21

Starkey, G. S. & McCandliss, B. D. (2014). The emergence of «groupitizing» in children's numerical cognition. *Journal of Experimental Child Psychology, 126*, 120–137. https://doi.org/10.1016/j.jecp.2014.03.006

Statistisk sentralbyrå. (2020). *Utdanning barnehager*. https://www.ssb.no/utdanning/barnehager/statistikk/barnehager

Stone, M. H. (1996). An essay on the ruler: What is measurement? *Rasch Measurement Transactions, 10*, 502–503.

Sundtjønn, T., Hølland, S., Bjørnestad, E. & Dalland, C. P. (2021). *Overgangspraksiser, læring og undervisningspraksiser – barnehage og førsteklasse En arbeidsrapport av litteraturgjennomgang*. https://www.udir.no/contentassets/af616ec30b0446fb-833024bcb151164a/sk-21-1-manus-lui-publisert.pdf

Sæbbe, P.-E. & Mosvold, R. (2020). The complexity of teaching mathematics in kindergarten: A case study and conceptualization. I *Mathematics education in the early years, results from the POEM4 conference 2018* (s. 385–400). https://doi.org/10.1007/978-3-030-34776-5_23

ten Braak, D. & Størksen, I. (2021). Psychometric properties of the Ani Banani math test. *European Journal of Developmental Psychology, 18*(4), 610–628. https://doi.org/10.1080/17405629.2021.1879046

Urban, M., Reikerås, E., Eidsvåg, Gunnar M. Guevara, J., Saegø, J. & Semmoloni, C. (2022). *Nordic approaches to evaluation and assessment in early childhood education and care* (TemaNord rapport 2022:512). http://dx.doi.org/10.6027/temanord2022-512

Vik, N. E. (2017). Språkkartlegging av flerspråklige barn i barnehagen-fra kontrovers til kompromiss? *Tidsskrift for nordisk barnehageforskning, 14*(10), 1–16. https://doi.org/10.7577/nbf.1734

Wager, A. A. & Parks, A. N. (2016). Assessing early number learning in play. *ZDM, 48*(7), 991–1002. https://doi.org/10.1007/s11858-016-0806-8

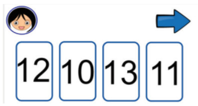
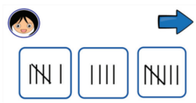
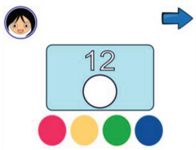
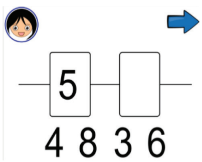
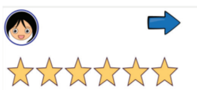
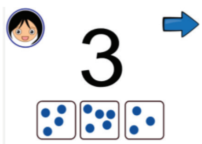
Whitacre, I., Henning, B. & Atabas, S. (2020). Disentangling the research literature on number sense: Three constructs, one name. *Review of Educational Research*, *90*(1). https://doi.org/10.3102/0034654319899706

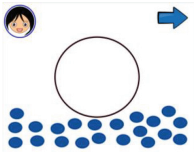
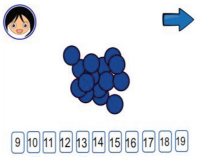
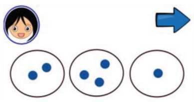
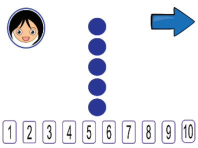
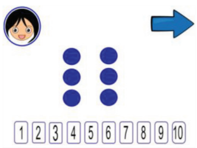
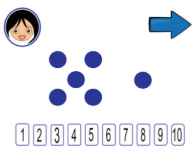
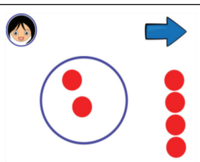
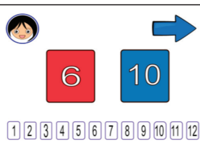
Wright, B. D. (1977). Solving measurement problems with the Rasch model. *Journal of Educational Measurement, 14*(2), 97–116.

Wright, B. & Linacre, J. (1994). Reasonable mean-square fit values. *Rasch Measurement Transactions, 8*(3), 370.

Appendiks 1

Eksempler på oppgaver fra hver FoNS-kategori.

Navn	Bilde	Instruksjon	Tallverdi
Tallidentifikasjon (TI)			
TI1		«Trykk på tretallet»	3
TI6		«Trykk på tallet 13»	13
TI13		«Trykk der du ser sju»	7
TI16		«Flytt riktig ball inn i hullet, trykk på ballene for å høre hvilket tall de har»	12
Systematisk telling (ST)			
ST1		«Flytt tallet som kommer etter fem inn i boksen.»	6
ST4		«Trykk på den andre stjernen i rekken.»	2
Tall og mengde (TM)			
TM1		«Trykk der du ser like mange baller som tallet.»	3

TM9		«Flytt fem baller inn i den store rundingen.»	5
TM13		«Flytt ballene for å telle dem, trykk på tallet som viser hvor mange baller det er.»	16
Mengdediskriminering (MD)			
MD1		«Trykk der det er flest.»	3
Subitisering (S)			
S4		«Hvor mange prikker så du?»	5 (rad)
S5		«Hvor mange prikker så du?»	6 (terning)
S11		«Hvor mange prikker så du?»	6
Aritmetisk kompetanse (AK)			
AK1		«Dra baller inn i rundingen slik at det blir fire til sammen.»	4
AK10		«Det er seks baller i den røde boksen. Det er ti baller i den blå boksen. Hvor mange flere er det i den blå?»	10 – 6 = 4

Paper IV

Validity evidence for the formative assessment of early number sense with digital tools and interactive items

Gunnhild Saksvik-Raanes and Trygve Solstad

NTNU - Norwegian University of Science and Technology, Faculty of Social and Educational Sciences,

Department of Teacher Education

Author note

This article is submitted to ESM.

The authors have no known conflict of interest to disclose.

The authors did not receive financial support from any organisation external to the NTNU for the submitted work.

Correspondence concerning this article should be addressed to Trygve Solstad, Department of Teacher Education, Faculty of Social and Educational Sciences, NTNU - Norwegian University of Science and Technology. Gunnerus gate 1, 7012 Trondheim.
Email: trygve.solstad@ntnu.no.

Acknowledgements

We would like to thank Tore Alexander Forbregd for invaluable contributions to the development of ENSA, Bjørn Evjen for contributions to the programming and scoring of items, Eivind Kaspersen for the support during the whole process from test development to data analysis and interpretations, Astrid Junker for contributing to the data collection and, together with Guri Nortvedt and Yvonne Grimeland, participating in the expert review groups during the item development. Finally, we thank Professor Jeremy Hodgen for critical review and suggestions in the writing process.

This paper is submitted for publication and is therefore not included.

Paper V

Factors influencing teachers' interpretations of assessment results

Gunnhild Saksvik-Raanes & Jeremy Hodgen

Abstract

This study investigates factors influencing teachers' interpretations of young children's number sense assessment results. Structured interviews were conducted with five teachers of first-grade children (aged 6-7) in Norway. The teachers were presented with the results of a specially designed instrument aligned with the curriculum and intended to inform their instructional decisions. A thematic analysis was conducted to identify the factors that influenced the teachers' interpretation of the assessments. Findings indicate that the instructional materials chosen by schools limited and shaped teachers' interpretations of children's assessment results; teaching objectives linked to the start of formal schooling reduced the formative value of assessments for teachers; and teachers' interpretations of children's performance were based on social and other non-cognitive factors. Together, these factors form a serious challenge to initiatives to enhance formative assessment by teachers. The implications for the design and use of effective diagnostic assessments, support materials and ways to enhance teachers' assessment literacy are discussed.

Keywords

Assessment literacy, formative assessment, instructional validity, number sense, primary school education

This paper will be submitted for publication and is therefore not included.

12.0 Appendix: Interview guide teachers

1. **Introduction**
   I will start the interviews with a short introduction about the project, ethical considerations, and conditions regarding the recording of the conversation. Each of the teachers will need to sign the consent form before we start the conversation. The interview will be conducted at the teacher's workspace after they finish their teaching duties.

2. **Questions related to children's number sense and teachers' expectations.**

   - From your experience, could you describe what children know about numbers when they start school?
   - What number sense activities do you usually include in your teaching in the first semester of first grade?
   - What should the children know regarding number sense at the end of first grade?

   After these initial questions, the teachers are introduced to results from a previous study where we found that children in Early years education demonstrated a number sense knowledge that exceeded the learning goals for the second grade in the Norwegian curriculum.

   - Do these results surprise you?
   - How would you consider these results compared to what you discussed previously?

3. **The teachers are shown an overview of their child's results and asked to discuss the following questions:**

   - Can you discuss what these results tell about your child's number sense?
   - Does this align with what you have seen from the children in the class? Is there anything that surprises you?
   - What do these results tell you about this child's number sense?
   - Can you say something about how you would plan for the further learning process for this child?

4. **Questions related to previous assessments and further work:**

   - What is your experience with using tests to describe children's competence?

- In general, how well do you think tests (like the National numeracy test) describe your children’s mathematical knowledge?
- What should a teacher consider when using tests to describe children’s knowledge?
- More specific about items? Reliability?

ISBN 978-82-326-7608-8 (printed ver.)
ISBN 978-82-326-7607-1 (electronic ver.)
ISSN 1503-8181 (printed ver.)
ISSN 2703-8084 (online ver.)

NTNU
Norwegian University of
Science and Technology