Number sense in young children—(inter)national group differences and an intervention programme for children with low and average performance

Helsinki 2006

Research Report 269

Pirjo Aunio

Number sense in young children—(inter)national group differences and an intervention programme for children with low and average performance

Academic Dissertation to be publicly discussed, by due permission of the Faculty of Behavioural Sciences at the University of Helsinki, in Lecture room 5, Main Building, Fabianinkatu 33, on April 8 th, at 10 o'clock

Helsinki 2006

Supervisors: *Professor*
Jarkko Hautamäki
University of Helsinki, Finland

Professors
Johannes E.H. Van Luit
Utrecht University, The Netherlands

Pre-inspectors: *Professor*
Erno Lehtinen
University of Turku

Professor Ernest
C. D. M. Van Lieshout
Vrije Universiteit Amsterdam

Custos:
Professor
Jarkko Hautamäki
University of Helsinki

Opponent:
Professor
A. J. J. M. Ruijssenaars
University of Groningen

ISBN 952-10-3006-2 (paperback)
ISBN 952-10-3007-0 (Pdf)
ISSN 1795-2158
Yliopistopaino
2006

University of Helsinki
Faculty of Behavioural Sciences
Department of Applied Sciences of Education
Research Report 269

Pirjo Aunio

Number sense in young children—(inter)national group differences and an intervention programme for children with low and average performance

---

Abstract

The purpose of this study was to develop a Finnish assessment tool for measuring young (4 to 7½ yrs.) children's numerical skills, to examine the number sense of children aged three to eight years with different national and performance backgrounds, and to investigate the effects of two mathematical intervention programmes. The theoretical framework was based on the development of specific and general numerical skills in early childhood. The two sets of skills, conceptualized here as relational and counting skills, were seen to develop in interaction, in which maturational (i.e. cognitive factors) and environmental (e.g., culture, motivation) factors play an important role.

The general research aims of the study were as follows: (1) to develop a number-sense screening tool for the use of Finnish early-childhood-(special) education professionals and for research purposes; (2) to compare the development of number sense in children with different backgrounds (age, gender, location/nation, language, performance); (3) to investigate whether it was possible to enhance the level of number-sense development in preschool years.

Five empirical studies addressed these questions: Study I examined the psychometric aspects of the Finnish Early Numeracy Test (the ENT-Fin), also producing the norms for Finnish children aged between four to seven-and-a half years; Study II examined the number sense of children from Beijing and Helsinki using the Chinese and Finnish ENTs; Study III focused on the number sense of children from Finland, Hong Kong, and Singapore, adding language (English vs. Chinese) as one controllable within-Asia variable in the cross-national research chain; Study IV was a preliminary investigation into low-performing children's number sense; and Study V investigated the effects of two mathematical-thinking intervention programmes. The effects of children's age and gender on number-sense performance were examined in studies I to IV.

The results of the empirical studies confirmed that the Finnish ENT (Van Luit, Van de Rijt, & Aunio, 2005) is a solid tool for screening the number sense of children aged four to seven-and-a-half years and is a sound assessment method for examining the numerical abilities of preschool children. The two-scale approach (relational and counting) was confirmed as valid. There were cross-national differences in that the children in Beijing, Hong Kong and Singapore had better number sense than the children in Finland. There was variation in performance within Asia, in that the children in Singapore outperformed their peers in Hong Kong. The Finnish children with special educational needs (SEN) and a multi-language background had lower number-sense performance than their average peers. The counting abilities in the multi-language group showed a different developmental trend than among the SEN children, in that the four-year-old multi-language children performed like SEN children but in the seven-year-old age group their performance was similar to that of children showing average development. The two intervention programmes, Let's Think! (Adey, Robertson, & Venville, 2001) and Count Too!

(Van Luit & Schopman, 1998), were found to increase children's specific numerical knowledge, however the between-group differences vanished when the interventions had finished.

Keywords*: cross-national differences, early numeracy test, general numerical skills, low performance, mathematical-thinking interventions, multi-language, number sense, special educational needs, specific numerical skills*

Helsingin yliopisto
Käyttäytymistieteellinen tiedekunta
Soveltavan kasvatustieteen laitos
Tutkimuksia 269

Pirjo Aunio

Lasten lukukäsite

Tutkimus kansainvälisistä ja kansallisista eroista lukukäsitteessä sekä matemaattisen ajattelun interventiosta

---

Tiivistelmä

Tutkimuksessa kehitettiin suomalainen lasten lukukäsitteen mittari, vertailtiin lukukäsitteen hallintaa eri kansalaisuuksilla, mitattiin suomalaisten erityislasten lukukäsitettä, sekä kokeiltiin kahta matemaattisen ajattelun interventiota suomalaisilla päiväkotilapsilla. Työn teoreettinen perusta on pienten (3–8 -vuotiaiden) lasten erityisten ja yleisten numeeristen taitojen kehityksessä. Nämä on operationalisoitu suhde- ja lukujonotaito käsitteinä. Suhde- ja lukujonotaidot kehittyvät vastavuoroisessa suhteessa, johon vaikuttavat merkitsevästi maturaatio (ilmenee mm. kognitiivisten komponenttien kehittymisenä) sekä ympäristötekijät (mm. kulttuuri, motivaatio).

Tutkimuksen tavoitteet olivat: (1) Kehittää suomalainen seulontamittari, jonka avulla varhais (erityis)opetuksen ammattilaiset kykenevät löytämään lapset, joilla on pulmia lukukäsitteessä; (2) Vertailla lukukäsitteen hallintaa erilaisen taustan omaavilla lapsilla – taustamuuttujina muun muassa asuinpaikka (Peking, Hong Kong, Singapore, Suomi), opetuksen kieli (englanti, kiina, suomi) ja erityisopetuksen tarve (oppimisvaikeus, monikielisyys); (3) Tutkia voidaanko 4–6 -vuotiaiden lasten matemaattisen ajattelun tasoa nostaa intervention avulla.

Tämä tutkimus koostuu viidestä empiirisestä osatutkimuksesta. Ensimmäisenä tutkittiin suomalaisen Lukukäsitetestin (Van Luit, Van de Rijt, & Aunio, 2005) psykometrisia ominaisuuksia ja tuotettiin normit mittamaan 4–7½ -vuotiaiden lasten lukukäsitettä. Toisessa osatutkimuksessa tarkasteltiin suomalaisen ja kiinan kielisen Lukukäsitetestin avulla lasten osaamista Helsingissä ja Pekingissä. Kolmas osatutkimus laajensi kansainvälistä vertailututkimusta mittaamalla lasten lukukäsitettä englannin, kiinan ja suomen kielisillä testeillä Hong Kongissa, Singaporessa ja Suomessa. Neljännessä osatutkimuksessa keskityttiin suomalaisten monikielisten ja erityislasten lukukäsitteeseen. Viidennessä osatutkimuksessa tarkasteltiin mahdollisuutta nostaa lasten (tavallinen tai heikko lukukäsite) matemaattisen ajattelun tasoa kahden interventio-ohjelman, Ajattele! (Let's Think!; Adey, Robertson, & Venville, 2001) ja Matikkaa! (Math!; Van Luit & Schopman, 1998), avulla.

Empiiriset tulokset osoittivat, että suomalainen Lukukäsitetesti on pätevä seula löytämään ne 4–7½ -vuotiaat lapset, joilla on pulmia matemaattisten esitaitojen kehityksessä. Tämän lisäksi mittaria ja sen kahta osatestiä voidaan käyttää tutkimuksiin, missä tarkastellaan 4–6 -vuotiaiden lasten matemaattista ajattelua. Kansainvälisissä vertailututkimuksissa havaittiin, että suomalaisten lasten lukukäsite oli heikompi kuin samanikäisten lasten Hong Kongissa, Pekingissä ja Singaporessa. Myös aasialaiset erosivat toisistaan, singaporelaisten lasten lukukäsite oli parempi kuin hongkongilaisten lasten. Suomalaisten monikielisten ja erityislasten matemaattiset taidot olivat heikommat kuin tavallisten lasten erojen näkyessä jo hyvin nuorilla lapsilla. Monikielisten lasten lukujonotaidot kehittyivät ennen koulua merkitsevästi saavuttaen tavallisten lasten osaamisen tason koulun alkaessa. Interventio tutkimuksen tulokset olivat ristiriitaisia, heti intervention loppumisen jälkeen opetusryhmän lasten erityiset numeeriset taidot olivat verrokkeja parem-

mat, mutta ero ei ollut merkitsevä puolen vuoden jälkeen. Ohjelmilla ei ollut vaikutusta yleisempiin matemaattisen ajattelun taitoihin.

Lyhyesti sanottuna nuorten lasten lukukäsitteen hallinnassa on tuntuvia kansainvälisiä eroja, joita selittävät muun muassa erot varhaisopetuksessa. Opetuksen ratkaisuin voidaan vaikuttaa lasten matemaattisen ajattelun kehitykseen ennen koulun alkua, mikä on oleellista etenkin kun kyseessä on erityislapsi tai monikielinen lapsi.

Avainsanat: *erityiset numeeriset taidot, erityislapset, kansainväliset erot, lukukäsite, Lukukäsitetesti, matemaattisen ajattelun interventio, matemaattiset esitaidot, monikielisyys, oppimisvaikeus, yleiset numeeriset taidot*

Acknowledgements

I wish to thank my supervisors professor Jarkko Hautamäki and professor Johannes Van Luit for their encouragement and support; my reviewers professor Erno Lehtinen and professor Ernest Van Lieshout for their valuable feedback; my licentiate thesis reviewers docent Jorma Kuusela and docent Raimo Lind for their encouraging critics; research fellow Markku Niemivirta and Pekka Heiskari of teaching me a lot of doing research and measurements in a proper high quality way; and researcher Pekka Räsänen for widening my knowledge about children's mathematical development. To make such a study could not have been possible without the open-mindedness of my Asian research partners, I wish to thank professor Zhang Meiling and professor Shi in Chinese Academy of Sciences for taking me in their academic group; and professor Jessie Ee and professor Audrey Lim for making it possible to continue in the cross-national research path.

The field work done in Finland was extensive, and it is not possible to thank all parties involved by name, I hope that all of the administrators, educators, children and parents accept my thanks in this form. I wish to thank my valuable contacts in the field of early childhood special education: Tero Laine, Riitta Tervo and Mervi Levanoja, as well as my important research assistants Laura Hautamäki, Saara Sajaniemi and Anna Berglind. My colleagues in the Special Education Section and especially Hanna-Maija Sinkkonen and Marja-Liisa Kieksi needs to be thanked for their various forms of support. Jukka Määttänen I thank for being in crucial help in the publication phase of this thesis. The study reported here was financially supported by the University of Helsinki Research Foundation.

I wish to thank my sister Marja for her special "sister role" in my life. Her husband Jaakko needs to be thanked of being in help in many practical problems occurring along the way and their sons, Joona and Mikko, being my "test children" for so many measurements. My in-laws Anita and Kari, as well as grand-grand mother Hilkka have been irreplaceable in their help with practical family issues. Most of all I owe my gratitude to my beloved husband Niko, without him I would not have started this project and certainly would not have completed it, and to my dearest children, Elmi and Ilpo, who make my life so much richer.

21 February 2006
Singapore

Contents

List of original publications *[Note 1]*

I Aunio, P., Hautamäki, J., Heiskari, P., & Van Luit, J. E. H. (in press). The Early Numeracy Test in Finnish: Children's Norms. Manuscript accepted for publication in the *Scandinavian Journal of Psychology*.

II Aunio, P., Niemivirta, M., Hautamäki, J., Van Luit, J. E. H., Shi, J., & Zhang, M. (in press). Young Children's Number Sense in China and Finland. Manuscript accepted for publication in the *Scandinavian Journal of Educational Research.*

III Aunio, P., Ee, J., Lim, A., Hautamäki, J., & Van Luit, J. E. H. (2004). Young children's number sense in Finland, Hong Kong and Singapore. *International Journal of Early Years Education, 12* (3), 195–216.

IV Aunio, P., Hautamäki, J., Sajaniemi, N., & Van Luit, J. E. H., Number sense in young low performance children. Manuscript submitted for publication.

V Aunio, P., Hautamäki, J., & Van Luit, J. E. H. (2005). Mathematical thinking intervention programmes for preschool children with normal and low number sense. *European Journal of Special Needs Education, 20* (2) 131–146.

1 Introduction

1.1 Number sense in young children

Many studies have been conducted over the years concerning international differences in school-age children's mathematical achievement. The Trends in International Mathematics and Science Study 2003 (Mullis, Martin, Gonzalez, & Chorostowski, 2004) for Grade 4 and 8 and the Program for International Student Assessment 2003 (PISA, 2003) for 15-year-old students are two examples. The number of countries participating in international comparisons concerning young children's mathematical knowledge is much lower (e.g., Miura, Okamoto, Kim, Steere, & Fayol, 1993; Stevenson, Lee, & Graham, 1993). In general, it seems that preschoolers' mathematical skills have not been widely studied (see e.g., Donlan, 1998; Geary, 1994; Nunes & Bryant 1996, for reviews). A lot of preliminary mathematical knowledge develops in early childhood, providing a basis for later formal mathematics learning, and also difficulties in mathematical development emerge early on. It is essential to acquire research knowledge on early number sense in order to understand the development and difficulties children encounter, as well as for planning the relevant educational support for these children. Just recently (see the special July/August 2005 issue of the Journal of Learning Disabilities, Vol 38) there have been long-awaited attempts to arouse research interest in and discussion on early identification and interventions for children with mathematical difficulties.

The study of young children's number sense requires an assessment tool. In this study such a tool was found in Dutch Utrecht's Early Numeracy Test (Van Luit, Van de Rijt, & Pennings, 1994), which has been translated and adapted to suit the Finnish context. The second focus in the research was on the general and specific mathematical skills of children from different locations or countries (Beijing, Helsinki, Hong Kong, Singapore and Finland) and with different abilities (average and low) aged between three and eight years. There have been many educational-support efforts developed in mathematics for school-age children. However, as the differences already begin to show in early childhood there is no reason why support for low-performing children should not also begin then. The third focus in this study was on the effectiveness of two intervention programmes, Let's Think! (Adey, Robertson, & Venville, 2001) and Young Children with Special Educational Needs Count Too! (Van Luit & Schopman, 1998), which were introduced to Finnish preschool children [Note 2]. The purpose was to investigate the possibility of enhancing the level of mathematical thinking of children with average and low number sense by exposing them to an enrichment programme in a preschool setting.

In brief, the focus in this study was on children aged three to eight years. It moved from measuring and describing average number-sense performance in different international locations to investigating the number sense of special-educational-needs children in Finland, and then developing mathematical-thinking interventions for Finnish children.

The present doctoral dissertation consists of a theoretical and methodological summary and the five original empirical studies. The summary section begins with a theoretical conceptualization of number sense, and the variations found in number-sense development are then discussed in a review of the findings related to international, national and gender differences. A description of the number-sense concept adopted in this study follows. The mathematical interventions and assessment tool used are presented, and the introductory chapter ends with a description of the current aims and methods. In the following chapters the set of original studies is overviewed with regard to the aims of the doctoral dissertation. The final chapter gives a general summary of the results and ends with a discussion of the theoretical and practical implications, and of the limitations and future challenges.

1.1.1 The psycho-educational view on mathematical thinking

Various concepts related to young children's early mathematical skills are referred to in the literature: basic number skills (Geary, 1994), concepts of numbers and counting (Fuson, 1988), informal mathematical knowledge (Ginsburg, Choi, Lopez, Netly, & Chi, 1997), number module (Butterworth, 1999), and number sense (Dehaene, 1997). Variation in terms is also apparent among researchers applying the same assessment method as in the study at hand: take, for example, preparatory and initial arithmetic (Van de Rijt & Van Luit, 1993), preparatory arithmetic skills (Schopman, Van Luit, & Van de Rijt, 1996), infant numeracy (Van de Rijt & Van Luit, 1999), early mathematical competence (Torbeyans et al., 2002; Van de Rijt & Van Luit, 1998), early mathematics (Aubrey, 2001), early numeracy (Aubrey, 1999; Van de Rijt et al., 2003; Van Luit & Schopman, 2000) and number sense (Schopman, Van Luit, & Van de Rijt, 1996). Despite the somewhat different emphases, all these terms refer to the numerical skills that children acquire, or have, before formal schooling, and which are essential for learning basic arithmetical skills (e.g., addition, subtraction, multiplication, and division) and for the development of future mathematical knowledge.

Bryant and Nunes (2002) suggested that the basis for children's mathematical development is logical thinking, the teaching of conventional counting systems, and a meaningful context for learning mathematics. According to the research on logical principles (see Smith, 2002), the development of mathematical thinking is related to children's growing abilities to understand and make relational statements (e.g., learning what it means that a number is equal to or more or less than another number). In other words, it concerns the ability to compare, classify and understand one-to-one correspondence and seriation. Being able to detect one-to-one correspondence and to seriate is essential for understanding cardinality and ordinality, which in turn is important for understanding number-word sequence (Bryant, 1996). The ability to numerically compare two sets is a vital aspect of conservation ability and related forms of numerical reasoning (e.g., Sophian, 1998), while the ability to classify is a fundamental element of mathematical reasoning in general (Smith, 2002).

The learning of a conventional counting system in early childhood begins with acquiring whole-number-word-sequence skills. Some authors consider these skills the basis of children's growing number awareness (Fuson, 1988; Gelman & Gallistel, 1978), which differs from the view put forward by Smith (2002). It is possible to distinguish six stages in the development of such skills: primary understanding of amounts, and acoustic, asynchronic, synchronic, resultative and shortened counting (e.g., Fuson, 1988; Van de Rijt, 1996). Primary understanding of amounts emerges at around two years of age when children show knowledge of how the different number-words refer to a different number of objects, but at this stage only very basic discrimination of amounts is possible. When they are at the acoustic counting stage, around the age of three, they can say number-words, but not in the correct order, and they do not necessarily begin with one: it is as if they are reciting a nursery rhyme. When they reach the asynchronic stage, around the age of four, they are able to say number-words in the correct order and to point to objects, but the words and pointing are not coherent. Six months later at the synchronic stage they are able to recite number-words and to mark the counted objects correctly, by pointing at or moving the objects, for instance. The resultative counting stage emerges around the age of five, when children are able to say number-words correctly starting with one, and understand that countable objects should be marked once and that the last said number-word indicates the number of objects in a set. Furthermore, they understand that number-words form a growing series in magnitude, meaning that the bigger one refers to a bigger amount. During the shortened counting stage, at around five-and-a-half years age, they are able to recognize the figure five, for instance, and can continue counting upwards from that. Thus their ability to operate with the number-word-sequence for whole numbers, and to use that in problem solving, increases substantially during these developmental shifts. For the Finnish readers it can be interesting to refer to Keranto's (1981) early research about the developmental connection between the mathematic-logical principles and counting abilities in preschool years.

The model of the central conceptual structure of children's conceptual development introduced by Case and his colleagues (e.g., Case & Okamoto, 1996; Griffin, 2003; Griffin & Case, 1998) represents one of the most recent and holistic attempts to define the development of mathematical knowledge. The part of the model that concerns the development of young children's general and specific numerical skills (cf. Case & Okamoto, 1996; Griffin & Case, 1998) focuses on the central concept of whole numbers. This is a cognitive structure that permits the child to interpret the world of quantity and numbers in increasingly sophisticated ways, to acquire new knowledge in this domain, and to solve the range of problems that it presents (Griffin, 2003). It develops through two stages in early childhood. During the predimensional period, at roughly four years of age, children have two separate mathematical schemas: the global quantity schema that permits them to answer questions about 'more' and 'less', and the initial counting schema that permits them to state 'how many objects are in a set'. The unidimensional stage, which emerges at around the age of six, is when the mental number line develops, and is the stage at which the two above-mentioned schemas are merged. The mental number line consists of knowledge of written numerals, knowledge of number words, the ability to point to objects when counting, and knowledge of cardinal set values. Within the central concept of number at each developmental stage, a reciprocal relationship exists between general (e.g., the categorization used in counting situations) and specific (e.g., knowledge of number words) numerical skills (Case, 1996). The development of number sense is thus a combination of progress in general and specific skills. In this process, associative and conceptual learning feed on each other in a reciprocal and dynamic way. Together these feedback loops form a hierarchical learning loop. When a child learns a specific skill, such as the ability to compare the numerosity of two sets, this learning has an impact on more general comparison skills. The same support mechanism works from general to specific skills as well. As a result of the activity within the learning loop, the rate of learning in low-exposure situations is faster than one might otherwise expect since it is mediated by general understanding and insights acquired in high-exposure situations.

The model developed by Case and his colleagues suggests that specific numerical skills are built on general numerical skills and are mostly affected by social, environmental and cultural factors. It seems that the development of general skills, in contrast, is more dependent on maturation than on direct environmental influence (e.g., teaching). Case also suggests that general skills may be influenced by factors other than maturation, such as instruction, which would be secondary in nature and occur through the mechanisms of a hierarchical learning loop.

This current research follows the theory of developmental cognitive psychology (Case & Okamoto, 1996; Piaget, 1965), according to which the mathematical thinking is an important and inter-dependent part of general thinking abilities. The development of mathematical thinking, including general and specific skills, is seen to occur in interaction with maturational and environmental factors. The term *number sense* is used because it refers directly to what is important: the child's understanding of quantity, number words and number symbols.

1.1.2 International differences in number sense

As preschoolers' number sense already reflects skills (e.g., understanding number words) that are influenced by the social environment, differences in mathematical performance between young children from different cultures are of interest. Cross-cultural comparisons of preschoolers' mathematical skills have shown that the mathematical performance of Asian children is better than that of their non-Asian peers: for example, young Chinese children consistently outperform their Western peers in abstract and object counting, concrete and mental addition and subtraction, and in the use of sophisticated strategies in mathematical problem solving (Ginsburg et al., 1997; Huntsinger, Jose, Liaw, & Ching, 1997; Miller, Smith, Zhu, & Zhang, 1995; Zhou, Cheng, Mottram, & Rosenblum, 1999; see Song & Ginsburg, 1987 for contrasting results). Differences in language, teaching and cultural ethos have been considered factors underlying Asian children's superior mathematical performance. These aspects are briefly discussed below.

Language. The nature and structure of the number-word sequence in the child's language is relevant in terms of their number-sense development in that it affects the emergence of counting abilities and the cognitive representation of numbers. There are differences between languages. For example, many European systems of number words are irregular, while the Asian systems, which are based on Chinese, are totally regular up to 100. The irregularity in many European languages means that the names for the words do not map onto the underlying base-10 structure of number systems. There is no obvious tens or unit value, and furthermore the unit value is sometimes spoken before the tens value, which makes number-word learning slow and tag errors common. There is a direct one-to-one relationship in Asian languages between the number words and the underlying base-10 values: eleven, for example, is ten-one (shi-yi).

English- and Finnish-speaking children learn number words to twenty largely as a rote sequence in which the words between ten and twenty are not related to the words below ten. A typical error that Finnish children make in trying to learn the rules for the number words beyond ten is to say "yhdeksäntoista, kymmenentoista" [nineteen, tenteen] (Kinnunen, Lehtinen & Vauras, 1994). This is identical with the error occurring in English-speaking children's number-sequence learning. The Asian number-word systems are easier to learn: it is enough to know the first nine words, the words for the powers of 10 (shi, bai, qian, etc.), and the order in which the words are said (from the largest value to the smallest). Chinese children make fewer errors in saying the words to 19 and learn the number-word sequence between 109 and 2000 earlier than English-speaking children in the United States (e.g., Fuson & Kwon, 1992). Of Finnish seven-year-old children, 63% can say number words correctly from 1 to 50 (Kinnunen et al., 1994), whereas almost all Chinese six-year-olds can say those from 1 to 100 (Yang et al., 1982). The number-word sequence also affects counting abilities (e.g., Nunes & Bryant, 1996): Chinese-speaking four-to-six-year-olds outperform their English-speaking peers in different object and abstract counting tasks, for instance.

The number-word sequence system also affects the conceptual understanding of numbers. For instance, Miura and her colleagues (Miura, 1987; Miura et al., 1993; Miura et al., 2000; Miura, Okamoto, Vlahovic-Stetic, Kim, & Han, 1999; Okamoto, Miura, Suomala, & Curtis, 1996) argue that the variability in mathematics performance is due to differences in a cognitive representation of number that is affected by numerical language characteristics differentiating Asian and non-Asian language groups. The organisation of the Asian numerical language assists children in developing cognitive number structures or representations that reflect the base-10 numeration system. Miura and her colleagues presented evidence that the number constructions of non-Asian-language first graders (American, Finnish, French and Swedish) differ from those of Asian-language speakers (Japanese and Korean): the former showed an initial preference for representing numbers with a collection of units, while the Asian children showed an initial preference for using a canonical base-10 construction to represent numbers concretely. Furthermore, the non-Asian children had less well-developed place-value understanding than their Asian peers.

Teaching. Cross-cultural studies on teaching in Asian and non-Asian schools (e.g., Perry, VanderStoep, & Yu, 1993; Stevenson & Lee, 1995; Stevenson et al., 1993; Stigler & Perry, 1990; Yoshida, Fernandez, & Stigler, 1993) have found differences in teaching methods (i.e. the amount, timing and purpose of seatwork, the use of manipulative material and word problems, the whole-class vs. small-group approach, the speed of instruction, requirements for higher-order-thinking schemata proceeding the lesson), in classroom organization, (i.e. organizing tables and chairs providing opportunities for child-to-child discussions, tolerance of off-seat behaviour), and the work of teachers (i.e. teaching as team-work, the amount of classroom responsibility, national curricula). Cross-national comparison studies have usually been conducted in primary and elementary schools in the United States of America, Japan and Mainland China.

The following internationally varying issues were addressed in this doctoral work. The first is the existence of national guidelines for teaching young children mathematics in Mainland China and Hong Kong, but the lack of it in Finland and Singapore. The second difference is the age at which children begin their primary education: it is six years in Hong Kong, Mainland China, and Singapore, and seven years in Finland. Thirdly, it is not custom to teach Finnish under-six-year-old children mathematical skills in organized and structured ways, whereas in Asian locations, Hong Kong, Mainland China and Singapore, the skills are taught from three to four years onwards. Furthermore preschool children in Finland are seldom expected to operate with numbers beyond 10, while in Asia it is an integral part of the curriculum in most preschools. As a result, not only do Asian children benefit from the logic inherent in their number-word-sequence system, they also have more opportunities to practice numerical skills.

Cultural ethos. Several studies (Campbell & Xue, 2001; Caplan, Choy, & Whitmore, 1992; Jose, Huntsinger, Huntsinger, & Liaw, 2000; Stevenson et al., 1993; Tuss & Zimmer, 1995) have explained the variation in mathematics performance in terms of the differing cultural values in Asian and non-Asian families, referring to values that emphasise learning, familial support and learning expectations. Li (2002, 2004) introduced the Chinese cultural model of learning, which divides Chinese beliefs about learning into four categories: (1) purposes, which emphasises lifelong moral self-perfection; (2) processes, which focus on learning the virtues of resolve, diligence, the endurance of hardship, perseverance, and concentration; (3) achievement standards, which aim at breadth and depth of knowledge, application, and the integration of knowledge and moral character; and (4) affect, which involves commitment, passion, respect, humility, and shame/guilt. Given this suggested profound importance of learning in Chinese cultures it is easy to understand the superiority of Chinese students in academic achievement. However, as Li (2004) points out, the cultural learning model is, at this stage, a theoretical model that needs to be substantiated with empirical data on children's performance in mathematics at different Asian locations.

Since cross-cultural comparisons of mathematical performance have, until now, mainly focused on specific skills, and particularly on children aged six years or older, knowledge of the cross-cultural differences in young children's combined general and specific numerical skills is limited. The language and learning culture are considered potential explanatory factors in the cross-national comparisons undertaken for the current study.

1.1.3 National differences in number sense

Just as it is likely that cross-national differences will emerge, individual variation within a nation is also to be expected. It was interesting to note that, although low number-sense performance in special-educational-needs (SEN) children has been taken for granted by researchers and educators, it was not easy to find research backing these claims. The comorbidity of mathematical and other learning difficulties has been demonstrated in many studies in which children with attention-deficit/hyperactivity disorder (ADHD), attention-deficit disorder (ADD) (e.g., DeShazo-Barry, Lyman, & Grofer-Klinger, 2002; Marshall, Hynd, Handwerk, & Hall, 1997) and language deficit (e.g., Arvedson, 2002; Fazio, 1999; Helland & Asbjornsen, 2003; Koponen, Mononen, & Räsänen, 2003) have been found to have problems with mathematics. So far, however, the research has concentrated on specific learning-difficulty groups and the mathematical difficulties of children aged six years and above.

The Finnish early-childhood-education policy also caters for children with special educational needs, which means that there is a wide variation in children's abilities. A number of children have diagnosed learning difficulties (7.6% of all preschool children in Helsinki; City of Helsinki, 2004), and they are entitled to special educational services. Furthermore, at least an equal number of children are thought to have developmental problems that will need special attention from early-childhood (special) educators. For foreign readers it is important to know that at this age the Finnish education system does not identify mathematical difficulties as a reason for special educational services. A new group of children has recently joined the Finnish early-childhood (special) education system, namely children with a multi-language background, usually due to their immigrant status. These children often need extra support to enable them to cope in the Finnish environment: their Finnish skills may be inadequate, for instance. The basic arithmetic skills of children with SEN and those whose first language is other than that used in the school have been studied in some extent (Demie, 2001; Sammons & Smees, 1998; Secada, 1991; Strand 1997, 1999), but as in the international research arena (see Kroesbergen & Van Luit (2003) for a meta-analysis), so in Finland is there a lack of research in the number sense of young children with difficulties in early-year development. In order to support mathematical development in early-childhood (special) education it is important to know what kind of number sense children have, and what difficulties they encounter.

A quite practical approach was taken in the fourth part of this study, as a heterogeneous low-performing population was taken as the target group (i.e. their performance was lower in a field of development other than number sense), and their number sense was checked. The aim was to see if they also lacked number-sense abilities. Of interest, too, were the intra-low-performance-group differences. Did the children with a multi-language background differ from another low-performing group, comprising children with neurological-based learning disabilities in their number-sense performance?

1.1.4 Gender differences in number sense

There are contradictory research results in children's mathematical performance and gender. Some research shows that girls and boys possess identical primary numerical abilities (e.g., Dehaene, 1997; Nunes & Bryant, 1996). The researchers analysing the British National Curriculum Key Stage 1 measurements (children aged four to seven years) have mostly reported girls outperforming the boys in basic arithmetic (Demie, 2001; Gorard, Rees, & Salisbury, 2001; Strand, 1997, 1999). Carr and Jessup (1997) report contradicting outcomes, as in their first school year, boys and girls may use different strategies for solving mathematical problems, but there is no difference in the level of performance. Gender differences in general and specific numerical skills at preschool age have attracted little attention (Torbeyns et al., 2002; Van de Rijt et al., 2003), and it would therefore be worthwhile to check such differences in young children's number sense. Based on previous research no gender differences are expected to emerge in young children's number sense.

1.1.5 The concept of number sense adopted in this study

Given the results of previous studies and theories of number sense in early childhood, it would appear that this study belongs to the research field falling within the oval in Figure1.

Biologically primary and secondary numerical skills. There are several theoretical approaches to the study of mathematics knowledge in early childhood (see Ruijssenaars, Van Luit, & Van Lieshout, 2004 for an excellent review). Taking the developmental-psychology view, Geary (1994, 2000) distinguishes two sets of numerical skills. The first comprise biologically primary qualitative skills (see also Butterworth, 1999; Dehaene, 1997; Wynn, 1998) such as understanding numerosity (e.g., determining the quantity of small sets of items by subitisation or estimation), ordinality (e.g., understanding the basic notion of more than and less than), counting (e.g., enumeration of sets up to three), and simple arithmetic (e.g., being aware of an increase/decrease in the quantity of small sets). These abilities are innate, and thus universal, and form a skeletal structure for later numerical development. The second set comprises biologically secondary number, counting and arithmetic competencies, which are culturally determined even if built on biologically primary systems (Geary 1994, 2000). Number and counting abilities refer to the knowledge of counting and the base-10 system, for instance, and the arithmetic competencies are needed in making computations and solving word problems. The distinction between the primary and secondary skills is that the development of secondary abilities varies from one culture or generation to the next, depending on school practices, for instance, while biologically primary skills have universal developmental patterns.

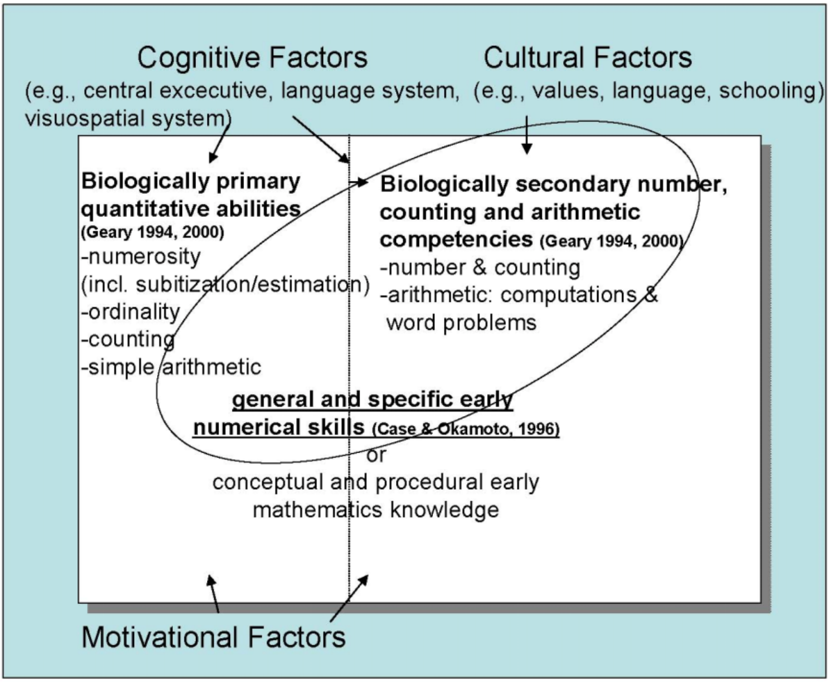

Figure 1. The theoretical perspective on young children's number sense adopted in this study

Situating the skills focused on in this study in the context of biologically primary and secondary numerical skills is not straightforward, as Geary defines all of the following as biologically primary: "The preverbal system does become integrated with the child's emerging language competencies (e.g., use of number words) although this system can still function without language. The result is verbal counting (e.g., counting items while starting 'one, two, three…') and the use of verbal counting to solve simple addition and subtraction problems, for instance, counting 'one, two…eight' to solve 5+3. By the end of the preschool years—even without formal instruction—children have a good, although not yet mature, understanding of counting concepts, they can use these counting skills to enumerate relatively large sets of objects and to aid in adding and subtracting objects from these sets, they have a basic understanding of ordinality (e.g., $1 < 2 < 3 <4$) and cardinality (i.e. that the last number word used while counting a set of objects represents the number of objects in the set), and can use these skills in very practical ways, as in measurement." (Geary 2000, II/12)

Biologically primary and secondary skills differ in the factors contributing to their development, culture having a bigger effect on secondary skills. However, as stated above, the primary skills are boosted at one stage by language development, which is in fact a cultural product. It is suggested here that such skills are no longer purely biologically primary, and that there is a transition period between the two sets of skills during which a mixture of cultural and maturational factors influence the emergence of the early number sense. The effects of maturation on development also decreases and the role of culture becomes more important. The focus in this current research is exactly on that transition phase of development, between the ages of three and eight.

Conceptual and procedural skills. As mentioned previously, number sense can be divided into two sets of skills, the general and the specific, as in the Case model, but another and more used approach is to divide the mathematical into conceptual and procedural abilities (e.g., Verschaffel, 2005). In general, conceptual skills refer to children's understanding of the logical principles needed in certain mathematical-problem-solving situations, such as knowledge about which strategies to use and why, and procedural skills involve the ability to use strategies correctly. In the context of this study, conceptual knowledge would refer to children's ability to organize and compare quantities (i.e. relational skills), and procedural knowledge to the ability to operate with number-word sequence (i.e. counting skills). This approach was not adopted because it offers no coherent view of the developmental relationships between the two sets of skills: there only seems to be agreement that conceptual understanding sometimes precedes and sometimes follows procedural abilities (e.g., Rittle-Johnson & Siegler, 1998). For the purposes of this study, the relationship between *general and specific numerical skills* in terms of the learning loop (Case, 1996) offers more potential for understanding the development of young children's number sense.

Cognitive and Motivational Factors. Number-sense development is closely linked to the development of general thinking skills, which in turn derive from functions in the other parts of the cognitive system such as the central executive (i.e. attentional and inhibitory control of information processing), and language and visuospatial systems representing and manipulating the information (Geary, 2004). The recent studies published by Hannula and her colleagues (e.g., Hannula, 2005; Hannula & Lehtinen, 2005) suggest that young children's Spontaneous Focusing on Numerosity, SFON, is one explanation for the development of early numerical skills in that children who spontaneously focus on the numerical information in their environment are more likely to acquire numerical abilities and as a result have better number sense than their peers whose focusing is less spontaneous. Individual motivation plays a vital role in the process of learning, as it is the essential explanatory concept that underlies learning and achievement behaviour such as choosing tasks, paying attention, expending effort, and showing persistence (Spinath, 2005). For instance, the lack of task motivation in mathematics has been found to contribute to slow skills development during the first three years of primary school (Nurmi & Aunola, 2005). The above-mentioned cognitive, attentional and motivational factors are relevant for the development of number sense, and especially important in explaining and understanding developmental problems. However, as the focus in this study is more on describing group differences than on developmental paths, these factors are not included in this research setting.

Cultural Factors. Differences in learning environments in terms of nationality (location) and language are one focus of this research, but there are also other contributory features (see the chapter on international differences in number sense). The cultural values that guide the instructional decisions are determined with reference to existing documentation (i.e. curricula) and relevant literature, and by working with local researchers in the field. The instructional-intervention study could be considered to be in line with cross-national comparison studies in that the learning environment is manipulated by providing children with additional learning experiences.

1.2 Mathematical-thinking interventions

The development of number sense is a combination of progress in general and specific numerical skills. In terms of learning and teaching, Case's (1996) view that the learning of specific numerical skills supports the learning of general numerical skills, and that this process is reciprocal, is accepted. Thus, ideal preschool instruction and good intervention combine these two aspects and provide settings that are interesting and motivating for children.

There is a wealth of literature on the teaching of various types of thinking, and there are several interventions covering specific school subjects for children with various skills and at different ages (e.g., Cooper, Charlton, Valentine, & Muhlenbruck, 2000; Hamers & Overtoom, 1997; Kroesbergen & Van Luit, 2003). It seems, however, that mathematical-thinking interventions have mostly targeted children of six years and above (e.g., Fuson, Carrol, & Drueck, 2000; Fuson, Smith, & Lo Cicero, 1997; Griffin & Case, 1998).

The study at hand used two programmes, 'Let's Think!' developed by Adey, et al. (2001), and 'Young Children with Special Educational Needs Count Too!' developed by Van Luit and Schopman (1998, from now on referred to as Count Too!). According to the classification developed by Hamers and Overtoom, (1997) both could be described as programmes with specific aims focusing on the thinking skills applied in one school subject, and both are based on the assumption that general thinking mechanisms can be influenced only through specific contents. Thus they also reflect the more general modes of cognition that are embedded in the content area of mathematics and science education.

Let's Think! (Adey et al., 2001; see Kuusela 2000 on successfully using cognitive acceleration programmes with Finnish upper-secondary-school pupils) relies on theoretical aspects, which together form a didactic model consisting of concrete preparation, cognitive conflict, metacognition, bridging, and (social) construction (e.g., Adey, 1997; Adey, Robertson, & Venville, 2002; Shayer & Adhami, 2003). Concrete preparation refers to the fact that the terms need to be established before particular problems are tackled. Cognitive conflict is based on the idea put forward by Piaget (1965) that thinking develops in response to cognitive challenge, hence experiences that are puzzling to a child and that cannot easily be explained using existing schemata may stimulate the development of more powerful schemata. Reflection on the process of problem solving is essential for the development of metacognition. This combines the ideas of Vygotsky and Piaget, and refers to the fact that cognitive development is helped if children are consciously aware of their own thinking, if they think of themselves as learners, and if they are continually made aware of their own active role in the learning process. Bridging signifies the fact that the reasoning patterns developed and applied in Let's Think! must be available in other contexts as well, and in this process the educator has a crucial role in demonstrating the relevance of learned reasoning patterns in other situations. Construction implies that children must create their own reasoning processes. The idea of social construction comes from Vygotsky (1978), who examined the process by which human beings grow up together watching and listening, trying things out in action and speech, looking for the effects on others, and so learning from each other. In Let's Think! each activities addresses one of the schemata of concrete operations, as follows: seriation, classification, time sequence, spatial perception, causality, and rules of a game (includes two schemata: theory of mind and concrete modeling).

Count Too! is based on several different psychological models that constitute the Dutch tradition of compiling well-known psycho-educational approaches (e.g., Van Luit & Schopman, 1998, 2000). The information-processing components refer to the models that are used in describing the use of knowledge, such as for perceiving, memorizing and retrieving information, and for using different strategies in solving mathematical problems. Cultural and historical psychology is represented through the theories developed by Galperin (1969) and Vygotsky (1978), and is evident in the emphasis on stages of the formation of mental actions and different types of orientation, the concept of the zone of proximal development, and the relations between instruction and development. Classical Gestalt theories are also identifiable (Wertheimer, 1971). All these psycho-educational elements refer to the use of concrete material in learning-development transitions, in combination with understanding the correct concepts, applying a variety of examples, utilizing group discussions in guiding the process of the formation of new mental actions, and focusing attention on the relations between cognition, emotion and motivation. The 20 Count Too! lessons focus on several age relevant mathematical-logical principles (e.g., classification, seriation, one-to-one correspondence) and basic counting abilities (e.g., synchronous counting, resultative counting, simple arithmetics) by using numbers 1 to 15.

Let's Think! and Count Too! share several characteristics. Firstly, they both aim to support the metacognitive abilities of children by focusing their attention on the ways in which they solve problems. Secondly, the idea of the zone of proximal development is embedded in them both, in that children are challenged by problems that are optimally difficult for them (Piaget, 1965), and the role of the teacher and of peers is to assist in finding the solution (Vygotsky, 1978). Thirdly, they require children to construct their own reasoning, a process in which the social group (Vygotsky, 1978) and the materials (Information processing and Gestalt theories) give relevant support. Fourthly, the programmes value the way of approaching problems, in that concrete preparation involving discussion and the use of materials (Galperin, 1969; Gestalt theories) are important prerequisites of the actual problem solving. Fifthly, they focus on transfer. Let's Think! emphasizes bridging through discussion, and Count Too! stimulates transfer by using diverse problems that show children how and when strategies can be applied in other situations. Finally, both programmes use one school subject as an instrument for forming thinking skills. The differences between the two are mainly qualitative: Let's Think! stresses general mathematical thinking skills more, and Count Too! specific mathematical-thinking skills, and cognitive conflict and bridging have a more pronounced role in Let's think!

The results of the study conducted by Adey et al. (2002) in which low-SES children were exposed to Let's Think! were positive, as the experimental group made significantly greater gains in cognitive development, measured according to their performance on Drawing and Conservation tasks (Piaget & Inhelder, 1974, 1976), over the intervention period than the controls did. When the boys and girls were investigated separately, the gains among the boys in the experimental group were not significantly greater than for those in the control group. The study carried out to investigate the effects of Count Too! on the number sense of children with special educational needs (mildly mentally retarded) revealed that those in the experimental group outperformed their controlled peers in the number sense test, but failed to transfer their knowledge to novel mathematical problems (Van Luit & Schopman, 2000), which is typical for children with mild mental retardation.

Here it is assumed that combining Let's Think! and Count Too! would generally accelerate young children's number sense, since the development of general mathematical-thinking abilities supports the development of specific mathematical-thinking skills and vice versa.

1.3 The assessment of number sense

Screening tests have been designed to differentiate those who are suspected of developmental delay from those who are developing normally. The main purpose of the Dutch Early Numeracy Test (ENT; Van Luit et al., 1994) is to identify children aged between four and seven years who are suspected of delay in terms of preparatory mathematics knowledge. Just as there is variation in the use of terminology referring to young children's number sense, the test has been given various names in English: it has been called The Utrecht Arithmetic Test for Toddlers (Van de Rijt & Van Luit, 1993), the Utrecht Test for Number Sense (Schopman et al., 1996; Schopman & Van Luit, 1995; Van Luit et al., 1994), the Utrecht Early Mathematical Competence Test

(Aguilar, Navarro, Alcade, Ruiz, & Marchena, 2005; Tzouriadou, Barbas, & Bonti, 2002), the Early Mathematical Competence Test (Van de Rijt & Van Luit, 1998) and the Early Numeracy Test (Torbeyns et al. 2002). The name Early Numeracy Test (the ENT) (in Finnish Lukukäsitetesti adapted by Van Luit, Van de Rijt, & Aunio, 2003/5, referred also the ENT-Fin) is used in this study. The test takes a developmental perspective on children's number sense, and aims at tapping eight aspects of numerical knowledge, including the concepts of comparison, classification, one-to-one correspondence, seriation, the use of number words, structured counting, resultative counting, and the general understanding of numbers.

The test is given individually and takes about 30 minutes for a child to complete. The items are scored by giving one point for a correct answer and zero for a wrong answer, the maximum number of points being 40 (e.g., Van de Rijt, Van Luit, & Pennings, 1999). The children are not given feedback as to whether their response is correct or incorrect. The test situation is not timed.

The original Dutch test (Van Luit et al., 1994) appeared in two parallel forms, A and B, while researchers doing longitudinal studies have also been using a third form, C, which is made up of particular items from the other two (e.g., Torbeyns et al., 2002; Van de Rijt et al., 2003). Van de Rijt et al. (1999) investigated the reliability and validity of the Dutch A and B forms, while simple Cronbach's alphas for the German, Flemish, Greek, Dutch, English and Slovenian A, B and C forms were presented in a European cross-national comparison study (Van de Rijt et al., 2003). Meanwhile Torbeyns et al. (2002) reported the reliability coefficients of the scores and the construct validity of the Dutch and Flemish A, B, and C forms. Only Form A is used in the current study, and its features investigated. The reason for not establishing the norms or using the ENT-Fin Form B, thereby also ruling out the use of Form C, was that, following a preliminary study in which the Finnish A and B forms were used, the ENT-Fin Form B was found not to be a real parallel with Form A, and it would have required a lot of effort to make it such. Consequently it was decided to focus on the ENT-Fin Form A.

Although the ENT is assumed to yield a common unidimensional measure of children's number sense (Van de Rijt et al., 1999), the practice of reporting the results runs contrary to this. For example, Schopman et al. (1996) reported two domains, mathematical prerequisites (i.e. items assessing concepts of comparison, classification, one-to-one correspondence, and seriation) and counting skills (i.e. items assessing the use of number words, structured counting, resultative counting, and general understanding of numbers), whereas Van Luit and Schopman (2000) reported three domains, math prerequisites, counting skills, and general knowledge of numbers (see Aguilar et al., 2005; Torbeyns, 1999; Tzouriadou et al., 2002 for yet another use of the test). From a theoretical point of view, the first four subscales of the instrument (concepts of comparison, classification, one-to-one correspondence, seriation) undoubtedly refer to the logical principles often identified as the key factors underlying children's understanding of quantities and relations (Piaget, 1965). Because of the type of skill these tasks measure (i.e. relational skills, or general numerical skills in Case et al.'s terminology), they are called relational tasks, and thus the relational scale. The rest of the test (the use of number words, structured counting, resultative counting, and general understanding of numbers) focuses more explicitly on the use and understanding of numbers (Fuson, 1988; Gelman & Gallistel, 1978), in other words counting skills, or specific numerical skills in Case et al.'s terminology, hence the term counting tasks and scale. The existence of the two scales in the ENT is investigated in this current set of studies across different samples. This it is not to suggest that using single-scale score is less relevant, but it is pointed out that the use of two scales may provide more information and a deeper insight into children's number sense.

The ENT has been analyzed in several studies, initially by Van de Rijt, et al. (1999) who measured the number sense of 823 Dutch children aged between four-and-a-half and seven-and-a-half. It has also been standardized for German children (Van Luit, Van de Rijt, & Hasemann, 2001). Neither of these studies considered the effect of the children's socio-economic background or gender on their number-sense performance. Flemish norms have been reported by Torbeyns (1999), who found no difference between the performance of Flemish boys and girls, but the effects of socio-economic factors were not assessed. Tzouriadou et al. (2002) studied the number sense of non-privileged and privileged Greek children. The results indicated that the privileged children had a better number sense than their non-privileged peers, however no results based on gender were reported. Aguilar et al. (2005) reported preliminary results from Spanish sample, but there was no reference to the background variables. Systematic research that takes children's background characteristics into account seems to be lacking.

One goal was for the ENT to establish a Finnish assessment tradition, as there is no standardized number-sense-screening test that can be used by early-childhood educators (see Korkman, Kirk, & Kemp, 1997; Wechsler, 1995 on tests for the use of psychologists), and that is also designed for children under six years old (see Ikäheimo, 1996; Vauras, Poskiparta, & Niemi, 1994 on measuring the abilities of children aged six and above).

1.4 Overall aims of the study

As the above literature review showed, early-number-sense development has not attracted wide interest: only limited cross-national comparisons have been made, and children with low performance are rarely studied from the number-sense-development perspective. Finland has no assessment tool that is standardised for measuring young children's number sense, and mathematical interventions have also been lacking. The study at hand was targeted to fill, at least partially, these regrettable gaps.

The core of this study is number sense in early childhood, which can be divided into general and specific skills. From this developmental-cognitive-psychological perspective, its development is seen to correlate highly with that of general thinking. The process starts in early childhood, hence age and the social and instructional environments are important contributory factors at an early age. The Early Numeracy Test (Van Luit et al., 1994; Van Luit et al., 2003/5), assessment tool used and studied here provides the tasks to measure early number sense, including general and specific mathematical skills. The mathematical-thinking intervention programmes target both of these aspects: Let's Think! (Adey et al., 2001) emphasises the growth in general (mathematical) thinking abilities, whereas Count Too! (Van Luit & Schopman, 1999) focuses more on specific mathematical skills. The study thus had three specific aims.

Aim 1. To develop a number-sense screening tool for the use of Finnish early-childhood-education professionals and for research purposes.

Aim 2. To compare the development of number sense in children (3–8 years) with different backgrounds (age, gender, location/nation, language, low performance).

Aim 3. To investigate whether it is possible to enhance the level of number-sense development in preschool years.

In line with these research aims, Study I investigated the psychometric aspects of the Finnish ENT, also producing norms for Finnish children aged between four and seven-and-a-half years; Study II examined the number sense of children from Beijing and Helsinki using the Chinese and Finnish versions of the ENT; Study III focused on the number sense of children from Finland, Hong Kong, and Singapore, adding the language (English vs. Chinese) as one controllable inside-Asia variable in the cross-national research chain; Study IV was a preliminary study on low-performing children's number sense; and Study V investigated the effects of two mathematical-thinking intervention programmes. The effects of children's age and gender on number-sense performance were considered in studies I to IV.

1.5 Overview of the methodological solutions

As the methods are reported in detail in the original publications, only a brief overview is provided here covering the studies conducted in 1999–2004 (Table 1).

Participants. A total of 1,995 children (1,079 boys and 916 girls) participated in these studies: 1,489 were from Finland (827 boys/662 girls), 376 from China (186 boys and 190 girls), and 130 from Singapore (66 boys and 64 girls). Overall the age of the children varied between 34 and 130 months (two years and ten months and 10 years and 10 months), the highest age being a result of the fact that some SEN children were exceptionally old: in general, the age varied between three and eight years.

Assessment. The main assessment tool used in these studies was the Early Numeracy Test (ENT) (Van Luit, et al., 1994; Van Luit et al., 2003/5), introduced in the previous section. The SRT I/ Spatial Relationship (Hautamäki, 1984; Shayer & Wylam, 1978), a shortened version of the Geometric Analogies (Hosenfeld, Van de Boom, & Resing, 1997) scales, and the Wechsler Preschool and Primary Scale of Intelligence-R (WPPSI-R) (Finnish edition, 1995) were also used in Study V.

Statistical methods. The studies followed three different sets of statistical solutions. Study I was unique in that the applicability of the new assessment tool was investigated by using the indexes of the Item Response Theory (Baker, 2001) and AN(C)OVAs.

One crucial methodological issue in this study, affecting Studies II, III and IV, was the fact that the ENT was used to measure and compare populations that had not been measured in that way before. This meant careful consideration of the measurement equivalency before the main analysis could be done (Van de Vijver & Leugn, 2000). Before we were able to analyse the measurement similarities in the first cross-national samples in Study II, we had to make a statistical decision. The fact that the ENT item scores are dichotomous (i.e. 0 or 1) resulted in some restrictions for choosing the analysis tool. Mplus (Muthén & Muthén, 1998–2004) was a statistical tool that made it possible to apply structural-equation modeling based on dichotomous variables, which was thus used in Studies II, III and IV. There was some variation in the statistical approaches selected: Study II used Multiple-group Means and Covariance Structures (MACS) analysis and Multiple Causes, Multiple Indicators Modelling (MIMIC), while Confirmatory Factor Analysis and AN(C)OVA with Bonferroni-adjusted multiple comparisons were the solutions adapted in Studies III and IV. Study V differed from the other studies on three major methodological issues: first, the pre-post-test (immediate and delayed) procedure was applied, as the other studies were based on one-time-measurements; secondly, assessment methods other than the ENT were used, and thirdly, alongside ANCOVA, the non-parametric Mann-Whitney U-test was used in the parts in which the sample size was small.

Table 1. Description of the methods used in the five original studies

Study	Year	Location/subjects	N (boys/girls)	Age	Main statistical methods	Purpose of the analysis
I	2002-3	Finland	1029 (550/479)	46-93	Descriptive statistics in age groups	Norms for Finnish children
					Cronbach's alphas	Reliability of the scores
					Item Characteristics Analysis	Analysis:
						item-characteristics curve,
						item difficulty and discrimination indexes,
						item correlation with the sum.
					ANOVA	Differences in the background variables
					AN(C)OVA (Bonferroni-adjusted multiple comparisons)	Effects of the background variables
II	1999	Beijing	130 (64/66)	55-90	Descriptive statistics	Investigation of too easy/difficult items
		Helsinki	203 (95/108)		MACS	Measurement equivalency
					MIMIC	Measurement equivalency and effects of nationality, gender and age
III	2002-3	Finland	254 (136-118)	45-96	Descriptive statistics	Investigation of too easy/difficult items
		Hong Kong	246 (122/124)		Confirmatory Factor Analysis	Measurement equivalency
		Singapore	130 (66/64)		Cronbach's alphas	Measurement equivalency
					AN(C)OVA (Bonferroni-adjusted multiple comparisons)	Measurement equivalency and effects of nationality, language, gender and age
IV	2002-4	Finnish		34-130	Descriptive statistics	Investigation of too easy/difficult items
		SEN	174 (130/44)		Confirmatory Factor Analysis	Measurement equivalency
		multilanguage	83 (52/31)		Cronbach's alphas	Measurement equivalency
		average performance	254 (132/122)		AN(C)OVA (Bonferroni-adjusted multiple comparisons)	Measurement equivalency and effects of age, gender, and group
V	2002-3	Finnish		56-79	Cronbach's alphas	Reliability of the scores
		average performance	33 (20/13)		Mann-Whitney U	Effects of the background variables and group
		low number sense	12 (7/5)		ANCOVA	Effects of the background variables and group

Note. Age is in months

2 Overview of the original studies

2.1 Study I

2.1.1 Aims

The aim of the study was to establish the norms for the Finnish Early Numeracy Test (the ENT-Fin), to estimate its reliability, and to assess the validity of the scores.

2.1.2 Participants and procedure

The participants were 1029 (550 boys and 479 girls) Finnish children ranging in age from 46 months (three years and 10 months) to 93 months (seven years and nine months). The children who had specific difficulties (i.e., special educational needs) that could have affected their test performance, varying from having a native language other than Finnish to facing problems that were neurological, were excluded from this study. Their performance was scrutinized in Study IV.

Eighty-nine administrators from various parts of Finland helped to collect the data, recruited mainly via the university teacher education departments. Most of them received face-to-face training in using the test. Special care was taken to ensure that only by using the test manual was it possible to conduct the measurement properly: for instance, a totally new chapter was written for the Finnish manual about the testing of young children's abilities, and pictures illustrated how to introduce the cubes to the children in tasks requiring the counting of organized and random sets of objects. The test administrators were also encouraged to contact the researcher if any problems arose. They were promised an ENT-Fin test manual free of charge when it was published for every 20 returned score sheets. Most of them were early-childhood educators measuring the number sense of the children in their teaching group. The test was administered to individual children in their own schools, usually in a separate quiet room with chairs and a table suitable for a child.

2.1.3 Measures

The Early Numeracy Test (Van Luit et al., 1994) was translated into Finnish in 1996. This translation was used in a study on 252 children aged between four years and seven months and seven years and six months (Kautonen, 1996), in which the item analysis and reliabilities were encouraging. The translation was checked in 2000, and consequently some small adaptations were made. The updated version was used in this study on a new set of data with a view to establishing Finnish norms.

There are 40 items in the test, measuring the concepts of comparison, classification, one-to-one correspondence, seriation, the use of number words, structured counting, resultative counting, and the general understanding of numbers. The items are scored by giving one point for a correct answer and zero for a wrong answer, the maximum number of points being 40 (e.g., Van de Rijt et al., 1999). The children are not given feedback as to whether their response is correct or incorrect.

The test was originally developed as a one-dimensional test, i.e. producing one overall score, but various researchers (Aguilar et al. 2005; Schopman et al., 1996; Torbeyns, 1999; Tzouriadou et al. 2002; Van Luit & Schopman, 2000) have developed different subscales. Given the theoretical construct and the results of Studies II and III it would appear to be reasonable to divide the ENT into two subscales, the relational scale (items 1–20) and the counting scale (items 21–40) [Note 3], for gathering information about young children's number sense. The applicability of the whole (items 1–40), including the relational and the counting scales, were scrutinized in this study.

2.1.4 Data analysis

The reliability of the ENT-Fin scores in this norm sample was assessed by using reliability coefficients (Cronbach's alphas) and item-characteristics analysis. The validity evidence of the norm-group children's test scores was based on analyses of the contents and the internal structure of the test, of the response processes for the various items, and of the effects of demographic variables (age, gender, mother's professional education, father's professional education, domicile, the number of children and the birth order in the family, the family form, and the child's hand preference) on the test scores (AERA, APA, NCME, 1999).

In order to establish the norms the children were divided into eight age groups at six-month intervals: four-year-olds (46–51 months), four-and-a-half-year-olds (52–57 months), five-year-olds (58–63 months), five-and-a-half-year-olds (64–69 months), six-year-olds (70–75 months), six-and-a-half-year-olds (76–81 months), seven-year-olds (82–87 months), and seven-and-a-half-year-olds (88–93 months). For the item-characteristics analysis they were divided into ten groups according to their performance level on the ENT-Fin whole scale. Although the test scores are age-related, the performance level gives more accurate results about the item's characteristics. The item-characteristics curve (e.g., Baker, 2001), the item-difficulty and discrimination indexes, and the item correlation with the sum were calculated for the analysis. Analyses of variance on the mean scores were used to measure the effects of the demographic variables.

2.1.5 Results

Finnish norms for children's performance in the ENT-Fin were established. By applying them it is possible to detect if the child is in need of extra attention in terms of number-sense development, for instance if his or her scores are below the minus-one standard-deviation line for the age group. In the oldest age group, with a mean age of seven-and-a-half years, the variation in the scores decreased, thus indicating some limitations in the use of the test. The age effect was clear and strong, as there was a steady increase in the mean raw scores on all three scales across the norm groups from four to seven-and-a-half years.

The Cronbach's alphas indicated good and acceptable reliabilities on all the scales in the whole sample, and for children under six-and-a-half years old. The reliabilities raised some concern in the over-seven groups on all of the scales. The item-characteristics curve showed the general trend: the better the performance level of the children, the more items they could answer correctly. The relational scale discriminated the low-performance children well, but was too easy for those with high performance. The counting scale was more challenging for all of the children. Meanwhile, the analysis based on the age groups showed that there were no items that were too easy for the children aged four and four-and-a-half. According to the item-difficulty index curve, there were ten items in the relational scale that were too easy for the children, and two items that had low correlations with the sum, one in the relational scale and one in the counting scale.

Evidence of validity was based on the content of the test and it was concluded that the original characteristics of the ENT-Dutch items had not been lost in the translation process. In terms of the response processes, there was evidence in five items of some confusion in the way that the children understood the question, while in terms of internal structure there were demonstrable developmental relationships between the relational and counting scales. The effects of the demographic variables showed that the mean raw scores increased in relation to age, as the younger children had lower scores than the older ones. There was also a gender effect favouring girls in the relational and whole scales. The effects of parental professional education were found to be significant: the children with more highly-educated parents achieved higher scores in the test, although there were some exceptions in the analyses scrutinizing the mother's and father's professional education separately, and in the children's performance on the relational and counting scales. The children living in small cities performed better than those in the metropolitan area, and the children from families with two to three children achieved better scores.

2.1.6 Discussion

The ENT-Fin is a good screening and research instrument for measuring the number sense of children aged four to seven-and-a-half years. However, since the variation in scores among those in the oldest group decreased, it is recommended that its use here is mainly for screening purposes. The items that were too easy, those with items with low correlations with the sum, and those that gave rise to some confusion in the response process carried clear implications in terms of the future development of the ENT-Fin test in that they required reformulation. The effects of age and parental professional education supported the claim that the Finnish ENT was a valid instrument in assessing children's number sense.

This study contributes in at least three ways to early-childhood research and practice. First, it produced the norms for Finnish children's performance in the number-sense test, unlike previous assessment practice aimed at early-childhood-education professionals. Secondly, the fact that the scales within the ENT have developmental relationships is new in applications of the test, and no such phenomena has been reported previously. It still needs to be determined whether the skills needed in relational tasks are crucial to, meaning strict prerequisites of successful performance on the counting scale. This would have some relevance in terms of instruction practices, for instance, in organizing the topics. Thirdly, a gender difference in young children's number-sense performance was found that was previously thought not to exist (e.g., Geary, 1994; Nunes & Bryant, 1996). The gender effect favouring girls was found to be bigger for the relational skills, possibly explained by the better pre-academic skills of girls (Hautamäki et al., 2002). However, as the effect was small and sample large, the result should be interpreted cautiously.

There were some limitations in this study. First, it was not possible to check the construct validity of the ENT-Fin by using another early-mathematical thinking test as no such test existed. Today the BANUCA (Räsänen, 2005) could be used, at least to some extent, for this purpose. Secondly, no test-retest procedure was applied. Thirdly, the power of the ENT-Fin to predict later mathematical performance was not examined in this study: this will be done later using longitudinal data collected for that purpose.

2.2 *Study II*

2.2.1 Aims

This study examined the influence of nationality, age and gender on preschoolers' number sense.

2.2.2 Participants and procedure

The participants were 130 Chinese children (64 boys and 66 girls) from two preschools and one primary school in Beijing, and 203 Finnish children (95 boys and 108 girls) from nine preschools and one primary school in Helsinki. The age of the children ranged from 55 months (four years and seven months) to 90 months (seven years and six months).

The data collection took place in 1999, when trained experimenters administered the test to individual children in their own schools, usually in a separate quiet room with chairs and a table. The experimenter provided the test materials (pictures, cubes, paper and pencil) according to the instructions for each item. Nine preschool teachers and one research assistant in Helsinki, and four psychology students in Beijing, conducted the test sessions.

2.2.3 Measures

The Early Numeracy Test (Van Luit et al., 1994) was used to assess children's number sense.

The draft of the Finnish ENT was used as it had been translated from Dutch into Finnish and tried out in a study with 252 children, thereby demonstrating the usefulness and psychometric adequacy of the Finnish scale (Kautonen, 1999). This was the first of the author's studies using the ENT, which was translated from Finnish and Dutch into Chinese for this purpose. The back-translation procedure was used to confirm the linguistic similarity of the test in both languages. Consultation with several experts in the field further ensured the cultural suitability.

2.2.4 Data analysis

As the ENT had not previously been used in Chinese, and more importantly, as the purpose was to compare the performance of children by using one instrument in different languages and cultures, it was crucial to ensure instrumental equivalency across the groups (Van de Vijver & Leugn, 2000). Multiple-group means and covariance structures (MACS) analysis is, in general, a good method for assessing measurement equivalence in different groups. However, a small sample size and having several groups to compare could make it less attractive. MACS was used in this study to investigate the measurement equivalency, thus when the focus was on the effects of nationality, gender and age in the latent factor the MIMIC (multiple causes, multiple indicators) modelling approach (cf., Muthén, 1989) was used. The MIMIC model allows for simultaneous confirmatory factor analysis and regression of the factor scores on several covariates without dividing the sample into numerous groups. The preliminary data screening was done by scrutinizing the descriptive statistics for the individual items within both nationalities and across genders. The purpose of this was to identify the items that were either too easy or too difficult and would decrease the quality of the data. The assumed latent structure underlying the children's ENT scores was then examined by conducting comparative MACS analyses in order to test the construct equivalence across the two nationalities. The next step was to apply the MIMIC modelling approach.

2.2.5 Results

Six items, all on the relational scale, were excluded from the analysis because they were too easy, resulting in a variation that was too small. The comparative MACS analysis revealed that the bi-dimensional structure, namely the relational and counting scales, underlies the children's test scores, and that a similar structure held for both the Chinese and the Finnish children. The direct effects of nationality on the latent factor in the MIMIC analysis showed significant mean differences favouring the Chinese children on both scales. No gender differences emerged in the latent means, but clear age effects were found. A significant nationality-by-age interaction effect was found on the relational scale, resulting from the fact that the Chinese children's level of relational skills as a function of age was increasingly higher than that of the Finnish children.

2.2.6 Discussion

The two-factor model (relational and counting scales) of children's number-sense was found and proved to be consistent in both the Chinese and the Finnish children's scores. This study was the first to analyze and report this kind of two-dimensionality. The relevance of this finding is that it increases the alternatives for studying number sense in general, and for using ENT scores in particular.

The Chinese children outperformed their Finnish peers on the relational and counting scales. As Chinese children begin to receive more structured mathematical teaching at preschool, and also begin primary school one year earlier than Finnish children, instructional policy is the most likely reason for these results. If we also consider the high prestige that good mathematical skills enjoy in Asian cultures, it is easy to understand the findings.

The boys and girls performed equally well in this study. Age had a significant impact on the test scores; the older children had better scores than the younger children. Both of these results are in line with previous research findings.

An interesting interaction effect was found in that the Chinese children improved their scores on the relational scale more than the Finnish children did. One possible explanation for this is that earlier exposure to mathematics teaching influences the relational knowledge of children. This supports the our theoretical standpoint concerning general and specific mathematical skills, and the other results reported in this doctoral thesis, that the origin and developmental paths may be different in the skills measured on the relational and counting scales.

This study had a few limitations. The samples were relatively small, and only collected in two capitals. Given the variation in school access and instructional level in the People's Republic of China, for instance, the results from the sample in Beijing cannot be generalized to the whole country. Nevertheless, as the comparisons were made between children from two capitals, the background factors in these samples were not necessarily very different. This comment leads to the second limitation, namely the lack of controllable background variables. In fact, questionnaires including questions about parental educational background and the home learning environment were also distributed to the parents. There was a 100% return in Beijing as 130 parents returned the questionnaire, and a 72% rate in Helsinki (109 questionnaires returned). Furthermore, nine teachers in Beijing and 28 in Helsinki filled in questionnaires about mathematical instruction. Analysis of these questionnaires would have given valuable information about the educational-cultural background of the children involved. However, it was impossible to use the information as there was no acceptable back-translation procedure applied in the design of the questionnaire, thus, at least to some extent, the different questions were asked in Finnish and in Chinese.

2.3 Study III

2.3.1 Aims

The purpose of this study, which followed Study II, was to examine the influence of age, gender, nationality, and language on children's number sense.

2.3.2 Participants and procedure

Two-hundred-and-fifty-four children in Finland (136 boys and 118 girls) from 33 preschools/schools, 246 children in Hong Kong (122 boys and 124 girls) from ten preschools/schools, and 130 children in Singapore (66 boys and 64 girls) from three preschools/schools participated in the study. The age of the children ranged from 45 months (three years and nine months) to 96 months (eight years).

The test was conducted in the children's language of instruction in the three locations. The children in Hong Kong were from schools where both English (n = 130; 66 boys and 64 girls) and Chinese (n= 116; 56 boys and 60 girls) were used as the language of instruction, and the test language was selected accordingly. The children in Singapore were attending schools in which English was the medium of instruction, and were thus also tested in English. Overall, Chinese was the test language for 116 children (56 boys and

60 girls) in Hong Kong, Finnish for 254 children (136 boys and 118 girls) in Finland, and English for 260 children (132 boys and 128 girls) in Hong Kong and Singapore.

The test procedure followed the same individual testing lines described in Study II. The data collection took place in 2002 when 71 early-childhood educators in Finland, nine research assistants in Hong Kong, and three research assistants in Singapore conducted the test sessions.

2.3.3 Measures

Chinese, English and Finnish versions of the Early Numeracy Test (Van Luit et al., 1994) were used to assess children's number sense. At this stage the original Dutch instrument had been translated into Finnish and its psychometrical properties tested in Study I. The English version was provided by the authors of the Dutch ENT, and the accuracy of the translation had been checked by British researchers in the field. The instrument was translated into Chinese for Study II. The cultural suitability of the instrument to all of the languages concerned was ensured by consulting several native-speaking experts in the field.

2.3.4 Data analysis

As in Study II preliminary data screening was carried out in order to identify the items that might disturb the analysis. Confirmatory Factor Analysis was then conducted in order to ensure construct comparability among the three languages. The Cronbach's alphas were used as indicators for internal consistency of the scales in the three languages. The effect of age on the scores was used as one indicator of measurement equivalence. The recommendations of Tabachnick and Fidell (1996) were followed in the main analysis, and separate ANOVA tests were used instead of MANOVA tests, as the two scales had high intercorrelations. If the ANOVA tests produced significant main or interaction effects, Bonferroni-adjusted multiple comparisons were then used. For one part of the analysis the children were divided into four age groups: four-year-olds (</= 57 months), five-year-olds (58–69 months), six-year-olds (70–81 months), and seven-year-olds (82=/< months).

2.3.5 Results

Following the preliminary data screening, five items belonging to the relational scale were excluded from the analysis as they were too easy for the children. Confirmatory Factor Analysis revealed that the two-factor structure (i.e. the relational and counting scales), underlies the children's test scores, and that a similar structure held for all of the samples. The internal consistencies of the scales in the three languages were good. The third aspect used to assess measurement equivalency was the influence of age on the scores: they were expected to increase with age, and such a trend was found except on the English relational scale. In this study gender was one variable in the main analysis. There were no statistically significant differences found between girls and boys number-sense performance.

The main analysis, in which the age of the children was controlled, revealed statistically significant differences in the mean scores, namely the nation effect was significant on the relational and counting scales. The Bonferroni-adjusted multiple comparisons showed that the Singaporean children outperformed the children from Finland and Hong Kong on both scales.

The analysis in which age was included as a categorical variable revealed statistically significant main and interaction effects, namely the main effects of age and nation were significant on the relational scale, as were the interaction effects of nation x age group and language x age group.

Both of these significant interaction effects came from unexpected low scores in two groups, the Singaporean seven-year-olds and the Chinese-speaking six-year-olds, which could have originated from weaknesses in the data sampling. The main effect of age group and nation were statistically significant on the counting scale, as was the interaction effect between them. The interaction effect was probably produced by the exceptionally low scores of seven-year-old Singaporean, again probably resulting from flaws in the sampling process.

The analysis of the counting scale conducted to investigate the language effects by comparing the Chinese-and English-speaking children in one Asian location, Hong Kong, revealed no statistically significant effects. However, the location had significant effects among the Asian English-speaking children as the Singaporean children outperformed the children in Hong Kong.

The analysis based on age-group division indicated that the children's performance on the counting scale according to nation was, on average, as expected: those from Singapore and Hong Kong achieved better means than those from Finland. Interpreting performance by age group on the relational tasks is less straightforward because there was not such a clear and repetitive rank order of scores among the three nations. The non-existence of a statistically significant effect of language on the counting scale in the different age groups was surprising.

2.3.6 Discussion

The purpose of this study was to continue and broaden the scope of the set of cross-national comparison studies involving Finnish and Asian children. The subjects were children aged between four and eight years from Finland, Hong Kong and Singapore, and they were tested in Chinese, English or Finnish. The hypotheses concerning measurement equivalence were supported in that the same two factors fitted all the language samples, and the age effects were also similar. In line with previous research findings exist (e.g., Geary, 1994; Nunes & Bryant, 1996), no gender differences were found in mathematical performance: the boys and girls achieved similar scores.

We then investigated whether the children in Asian cities had more highly developed numerical skills than those in Finland, and we found that the children in Singapore and Hong Kong did outperform those in Finland in terms of number sense. There are at least two plausible explanations for these differences: one is the one-year earlier start of primary school in Hong Kong and Singapore, and the other concerns the quality of numerical instruction in Asian locations. The latter in particular is considered to originate from the fact that the culture places high value on mathematical knowledge.

Contrary to expectations based on previous research results suggesting that the Chinese language is efficient in terms of supporting the early learning of mathematics (Fuson & Kwon, 1992; Ginsburg et al., 1997; Miller et al., 1995; Miura et al., 1993; Zhou & Boehm, 2001), no language differences were found in the counting-scale scores. Due to the lack of information about the children's home language in this study, replication of these measurements including the home and school language as a factor is suggested.

There were a few limitations. The major one was in the selection of subjects, as the sample was not randomly collected, and even if the total number of subjects was acceptable for a comparative study, it was rather small for testing simultaneously the effects of age, nation and language. The lack of controllable background variables was another limitation here as it was in Study II. Further studies in these locations using data on the home-learning environment, including the language factor would provide more holistic knowledge of number-sense development in young children. We might have produced a neat research continuum if the analysis had been done analogously with Study II, but this was not possible at the time of reporting the results. The data will be re-analysed, and the results reported if interesting, and especially if contrasting issues emerge. This further research will merge the Chinese and Finnish children's scores from Study II with the data used in this study so as to enable the effects of language on ENT performance to be assessed.

2.4 *Study IV*

2.4.1 Aims

The aim of this study was to investigate the number sense of young low-performing children.

2.4.2 Participants and procedure

The participants were 174 (130 boys and 44 girls) children with special educational needs (SEN), 83 (52 boys and 31 girls) children with a multi-language background, and 254 (132 boys and 122 girls) children with average development. The age of the children varied from 34 months (two years and 10 months) to 130 months (10 years and 10 months).

This data was collected partly in connection with the collection of the norm data for Study I. The test administrators alerted the researcher by memo if the child had previously shown developmental features that might affect his or her performance in number-sense testing: these children are henceforth referred to as low performers. Additional data was collected by university students specialising in special education. The data collection took place in 2002–2004.

2.4.3 Measures

The Early Numeracy Test (Van Luit et al., 1994; Van Luit et al., 2003) was used to measure the children's number sense. Two scales, the relational and the counting scale, were used in the analysis.

2.4.4 Data analysis

As in Studies II and III, preliminary data screening with verified the quality of the data, and then a Confirmatory Factor Analysis was conducted in order to ensure construct comparability among the samples. The effect of age on the scores was used as one indicator of measurement equivalence, and Cronbach's alphas were used as indicators of internal consistency in the group scales. The non-existence of gender differences was used as one indicator of measurement validity. Separate ANOVA tests were used, as the two scales had high intercorrelations. If the ANOVA produced significant main or interaction effects, it was followed by Bonferroni-adjusted multiple comparisons. For one part of the analysis the children were divided into three age groups: there were 113 children aged five and below (</= 69 months), 265 six-year-olds (70–81 months), and 133 aged seven and above (82 =/<).

2.4.5 Results

The same two-factor model fitted the data better than the one-factor model in all of the samples, namely in the whole sample, among all the girls and all the boys, and in the reference, multi-language and SEN samples. The use of separate relational and counting scales in the analysis was then ratified. In general, there were significant gains in all of the samples' on both scale scores according to age, with the exception of the relational scale in the multi-language group. The Cronbach's alphas as indicators of internal consistency of the scores in the samples were good on average. In general there were no gender differences, except that in the whole sample the girls performed better on the relational scale.

In the main analysis the reference group outscored both of the low-performance groups—the multi-language and SEN children—on the relational and counting scales. This trend was consistent across all three age groups, except that in oldest group the multi-language children performed as well as the reference children. There were two group differences in performance between the multi-language and the SEN children, namely on the counting scale in the sample including all children, and in the oldest age group the multi-language children achieved higher scores than the SEN children. The hypothesis that the counting-scale performance of the multi-language group would be similar to that of the reference group was not supported, as this was the case only in the oldest age group.

2.4.6 Discussion

The aim of this study was to investigate the number sense of low-performing children at an average age of six years and four months. The children were divided into three groups; the multi-language group consisting of children who had a language other than Finnish as their home language, the special-educational-needs (SEN) group in which the children had diagnosed or pre-assumed learning difficulties, and the reference group of children with average development.

The preliminary analysis proved the measurement equivalences across the samples, and allowed the group comparisons to be made. One gender difference was found, namely in the sample consisting of all of the children (N=511), the girls had better relational skills than the boys. Previous research has established that there are no gender differences in young children's mathematical skills (e.g., Carr & Jessup, 1997; Dehaene, 1997; Nunes & Bryant, 1996), so the current results are clearly in contradiction. As a similar result was also found in Study I and in basic arithmetics of same age children (Demie, 2001; Gorard et al., 2001; Strand, 1997, 1999), it would be worthwhile to focus on gender differences in future studies. However it should be noted that the differences occurred when the samples were large, in excess of 500.

The low-performing children were shown to have weaker number sense than those in the reference group. This is a significant result in terms of designing early-childhood (special) education programmes as support for the development of mathematical thinking should also be provided for young low-performing children. In general the number-sense performance of the multi-language and the SEN children was quite similar, but different in counting skills, which indicates that they probably need different kinds of instructional support. It is suggested that children with a multi-language background should be provided with Finnish-language enrichment on a large scale, incorporating mathematically relevant concepts, in their early-childhood (special) education. Cognitive-acceleration programmes such as Let's Think! (Adey et al. 2001) could be more useful for the SEN children as they probably also need support in terms of their general thinking abilities. However, we need more research to find the most beneficial (special) education support for these children.

2.5 Study V

2.5.1 Aims

The purpose of this study was to investigate the possibility of enhancing the level of preschoolers' number sense by introducing two intervention programmes, Let's Think! (Adey et al., 2001) and Young Children with Special Educational Needs Count Too! (Count Too!; Van Luit & Schopman, 1998). [Note 4]

2.5.2 Participants and procedure

The participants of the study were 45 (27 boys and 18 girls) Finnish children aged between 56 months (four years and eight months) and 79 months (six years and seven months). The children were from six preschools. The experimental group comprised 22 children and the control group 23 children. There were 12 (seven boys and five girls) low-number-sense performers, five in the experimental group and seven in the control group. The rest of the children had average number sense.

The instruction in the experimental condition was given to four small groups of five-to-six children. All of these groups followed both programmes, Let's Think! (Adey et al. 2001) and Count Too! (Van Luit & Schopman, 1998). There were two sessions in a week during the academic year 2002–03, approximately 60 sessions in all. The children in the control group followed average preschool activities.

2.5.3 Measures

The Early Numeracy Test (Van Luit et al., 1994; Van Luit et al., 2003) was used to measure the children's number sense. The relational and counting scales were used in the analysis. Spatial awareness and analogical reasoning were measured on the SRT I/ Spatial Relationship (Shayer & Wylam, 1978) and a shortened version of the Geometric Analogies (Hosenfeld et al., 1997) scales. The Wechsler Preschool and Primary Scale of Intelligence-R (WPPSI-R) (Finnish edition, 1995) was used to measure the children's general thinking abilities. A pretest – intervention – post-test design was applied. The ENT, the SRT I/Spatial Relationship scale and the Geometrical Analogies scale were used three times: pretest, immediate post-test, and follow-up (six months after the intervention). The WPPSI-R was used once, at the beginning of the intervention phase.

In the original publication there is one terminology mistake which needs to be corrected, namely the SRT I/Spatial relationship score is *not near but far transfer* indicator. Far transfer tasks involve skills and knowledge being applied in new chancing situations, as in near transfer tasks the skills and knowledge are applied the same way every time. The SRT I/spatial relationships tasks are clearly far transfer tasks in relation to skills practiced in Let's Think! and Count Too!

2.5.4 Data analysis

The internal consistency of the scores on the scales at the first measurement time is described in terms of Cronbach's alphas. The groups were then compared at each measurement points, and the gain scores on each scale, namely between the immediate post-test and the pretest (Gain 1) and between the follow-up test and the pretest (Gain 2), were computed.

2.5.5 Results

The internal consistencies of the children's scores on the scales were acceptable. The main analysis, in which IQ was controlled, showed that the children in the experimental group had higher performance on the relational and counting scales in the immediate post-test than the children in the control group. The same between-group effect was found in the gain scores. There were no between-group differences at the time of the follow-up test.

The preliminary analysis revealed that the low-performing children in the experimental group had better means on the counting scale in the immediate post-test, and on the SRT I/Spatial relation scale in the follow-up test, than their peers in the control group.

2.5.6 Discussion

The research aim was to investigate whether it was possible to raise the level of young children's number sense by applying two intervention programmes, Let's Think! (Adey et al. 2001) and Count Too! (Van Luit & Schopman, 1998). The results of the study were conflicting. They were positive at the time of the immediate post-test in that, on the group level, the experimental group showed better development, especially in specific mathematical tasks: relational and counting skills were better in this group than in the control group. The results were less encouraging six months after the intervention, however, when the children in both groups performed equally well. It seems that it was possible to teach the children some specific mathematical skills, but there was no larger effect on the general level. These findings are comparable with those of other studies in which low-performance children do not show generalization of learned knowledge to other domains (Van Luit & Schopman, 2000), and consequently over time they need additional support in their mathematics development. The results with low-performing children were encouraging in this study, but need to be replicated with a larger sample.

The major limitation in the study was the small sample size, which restricted the use of analysis tools, for instance. Moreover, the simultaneous implementation of two programmes limited the conclusion possibilities, as it was impossible to determine whether a single programme would have produced the same results. The relevance in terms of future interventions is that, firstly, there are possibilities of supporting the development of number sense among normal-developing and low-performing children. Secondly, the interventions should be conducted by the educators in their own teaching groups, and the educators should have professional training in using the programmes and continuous consultation possibilities during the process. The future aim is to continue the research in order to find the most efficient interventions for low-performance children.

3 General summary and discussion

The aims of the present doctoral dissertation were, first, to develop a Finnish assessment tool for measuring young children's number sense, then to investigate the number sense of children with various national and performance backgrounds, and finally to test the efficiency of two mathematical-thinking intervention programmes, Let's Think! (Adey et al., 2001), and Count Too! (originally called Young Children with Special Educational Needs Count Too! by Van Luit & Schopman, 1998).

3.1 Summary of the main findings

The results showed that it was possible to develop a screening instrument for Finnish early-childhood (special) educators for measuring the number sense of young children: the Finnish Early Numeracy Test (Van Luit et al., 2005) proved to be a reliable screening tool for children aged four to seven-and-a-half years. The two-scale (relational and counting scales) alternative to the one-scale possibility (ENT whole scale) introduced by the original test developers proved to be a valid and useful research approach to measuring young children's number sense.

In the cross-national comparisons the children in the Asian locations, namely Beijing, Hong Kong and Singapore, outperformed the Finnish children in number-sense tasks. There were also differences found between the Asian locations, namely the children in Singapore outperformed their peers in Hong Kong. Instructional decisions that are the most likely explanatory factors here, as primary school starts one year earlier in Asian locations than in Finland, for instance, and early mathematics is much more prominent in the early-childhood instructional practices. The cultural tradition of placing high value on mathematics learning at an early age supports this reasoning.

The analysis of the number sense of Finnish low-performing children showed that the relational skills in the multi-language and SEN children were similar, but some between-group differences existed in counting performance, favouring the multi-language group. Both low-performance groups (multi-language and SEN) were weaker than the reference group in this respect even among the youngest children. It was concluded that the developmental problem that caused the test administrators to single out these children as low performers also influenced their number-sense development. Furthermore, having more than one language in use at home and at preschool, affected the development of number sense in preschool years.

The results of the intervention study were somewhat encouraging, as the experimental group had better number-sense performance immediately after the intervention. It was disappointing that the effect did not persist in the follow-up measurement six months later, and there was no transfer effect in either of the post-intervention measurements. It appears to be easier to teach children specific numerical knowledge, but harder to influence their general thinking capacity.

The effect of gender was measured in Studies I to IV, but only in those using the extensive (> 500 children) Finnish data (I and IV) was the effect of gender statistically significant, as the girls outperformed the boys.

3.2 Theoretical implications

One clear benefit of this research is that it has demonstrated across the samples that it is possible to distinguish two sets of tasks in the ENT, relational and counting tasks. The two-scale approach enable young children's number sense to be investigated from a different angle than with the one-scale approach (i.e. total score). The fact that the scales showed high intercorrelations indicates that they are not totally separate sets of skills. It was shown that it was possible to use two scales, but this does not rule out the possibility of using one total score, which may be useful for certain purposes.

Another theoretically interesting issue arose when attempts were made to combine Case's model of the central structure of numbers (Case & Okamoto, 1996) and an existing assessment tool, the ENT, which in fact was originally developed from a slightly different theoretical perspective combining the then separate ideas of logical thinking (Smith, 2002) and number-sequence skills (Fuson, 1988) as the core of number sense. As a result, an interesting controversy arose: in Study II the relational tasks were shown to be in line with Case's general thinking skills, and the counting tasks to resemble specific mathematical skills, yet relational and counting tasks were used as indicators of the children's specific mathematical-thinking skills in Study V. This controversy is, in fact, not a controversy at all, and merely indicates how relative the line between the set of skills involved in number sense are. The counting scale quite clearly includes tasks that require children to use specific numerical knowledge, while the relational tasks are not so focused on numerical knowledge and are more concerned with measuring the ability to understand the relations between the aspects introduced in them. Thus the relational scale could be considered less numerical, and therefore analogous to general mathematical skills. Nevertheless, since it involves numerical information, it could in certain contexts be regarded as a measurement of specific numerical abilities. This context was present in Study V, in which the other instruments measured children's understanding of spatial relationships (SRT I/ Spatial Relationship, Shayer & Wylam, 1978) and analogical thinking (Geometric Analogies, Hosenfeld et al., 1997), which could be equated to Case's general mathematical skills. The model of the central structure of numbers is a tempting theoretical basis for future studies, as one of particular interest could be the reciprocal development of specific and general numerical skills. It would have been consistent with the theoretical construct adopted in this study to use the assessment tool developed by Case and Griffin, the Number Knowledge Test (e.g., Griffin, 2003), which measures the mathematical thinking abilities of children aged four to 10 years. It was not used because it was not published in time, but it has been used it in some preliminary studies (e.g., Kaminen, 2004), and it seems to be tempting research tool in the Finnish context.

The third theoretically interesting outcome was the girls outperforming the boys in Study I and IV. The results are inline with the research focusing on the same age children and little different mathematical skills (Demie, 2001; Gorard et al., 2001; Strand, 1997, 1999), but contradict the research results from basic numerical skills (e.g., Dehaene, 1997; Nunes & Bryant, 1996). We deem that the controversies about the gender effect are due to, on one hand, the different mathematical skills measured, and on other hand, the amount of controllable background variables included in the analysis. The future research is suggested as at the preschool age children shift from using the biologically primary quantitative abilities to learning the biologically secondary number, counting and arithmetic competencies, the latter being affected more by learning environment than the former (Geary, 1994, 2000).

3.3 Practical implications

One practically relevant outcome of this study is the Finnish Early Numeracy Test (Lukukäsitetesti, Van Luit et al., 2005), which enables professionals in early-childhood (special) education to measure the number sense of children aged between four and seven-and-a-half years. The main purpose of the test is to screen the children who have problems in their number-sense development. It is also useful for research purposes, as it distinguishes well the number sense of children between the ages of four and six on all performance levels.

A second practically interesting finding was the weak number sense of children with SEN and a multi-language background. They clearly need educational support already in preschool, it is thus necessary to continue the research to find the most suitable (special) education support for them. The children with a multi-language background seem to benefit from the general preschool education, as in the youngest group their performance was similar to that of the SEN group and in the oldest group they performed as well as the reference children. The similar phenomenon have been reported also in the Key Stage 1 studies (e.g., Strand, 1997, 1999).

A third practically important outcome was that indeed turned out to be possible to enhance the level of mathematical performance in preschools by providing well-planned instruction. In our cross-national studies we suggest that the better number-sense performance in Asian children are explained by the one-year earlier start of primary school in China and Singapore. The other explanation concerns the quality of numerical instruction in Asian locations (Ginsburg et al., 1997; Perry et al., 1993; Stevenson et al., 1993), which is considered to originate from the fact that the culture places high value on mathematical learning and knowledge (e.g., Huntsinger et al., 1997; Li, 2002, 2004; Sharpe, 2002). The results of the studies conducted in Asian locations support statements about the need for and benefit of good mathematical instruction in preschools. However, the word ‘good’ should be emphasised, as it refers to the educator’s good knowledge of mathematical thinking, so that he or she can choose the tasks that really enhance the level of mathematical thinking and support its development in children. This means a holistic approach to learning mathematics at preschool, such as the Realistic Mathematics approach (Ruijssenaars et al., 2004) applied largely in the Netherlands. Furthermore, educators should be aware of the knowledge levels of learners to be able to adjust the instruction accordingly. To be realistic, it will take time before early-childhood educators have such a repertoire of skills, and it will also require a critical review of Finnish teacher education in the field of early-mathematics teaching. An alternative and probably more immediate solution would be to provide educators with professional training and the possibility to use separately developed intervention programmes for supporting children’s mathematical thinking: the Count Too! (Van Luit & Schopman, 1998) and Let’s Think! (Adey et al., 2001) programmes adopted in this study are good examples, although there are others, such as Number Worlds

(Griffin & Case, 1995; Griffin, 2000). The experience gained in Study V will guide future interventions in that, firstly, there has to be training for educators in the use of the instruments, as well as ongoing consultation sessions during the implementation process to enable the educators to discuss and solve the problems that arise in the instruction. There should also be pre- and post-assessments of the children's abilities to facilitate assessment of the usefulness of the instructional action.

The fourth practical issue is the fact that the cross-national comparison results can be interpreted also differently: they can be seen to suggest that there is no need for early exposure for mathematics as the children's results at age of seven are quite similar even though the Finnish children got very little mathematic exposure in preschool years. This statement cumulate from the fact that Finnish children's results at age of seven are at level of their Asian peers even though the mathematics teaching begins later in Finland than in Asian locations. Yet this study cannot validate such a conclusion, the similar performance can be due to the ceiling effect of the test or sampling problems. To be able to make such a conclusion we need a longitudinal study with more appropriate assessments tools for primary grade children.

3.4 Limitations and future challenges

There are certain limitations in the studies that comprise this doctoral dissertation, one of which concerns the restricting of the focus to the number sense constructed in the ENT, and the fact that only one measurement instrument was used in most of the studies. It is likely that this focus left some important number-sense skills or areas untouched.

Secondly, the generalization power of the cross-national comparisons was low, as there was no randomised data sampling. However, as the results of both of the cross-national studies were coherent, it is tempting to believe that the phenomenon found really exists, namely that Finnish preschool children are lagging behind their Asian peers in terms of number sense. The investigation of cross-national differences would have been more profound if quantitative and qualitative data had been used for explaining the phenomenon.

Thirdly, introducing two different intervention programmes at the same time to the same children made it impossible to confirm the effects of either one. Nevertheless, since the study conducted by Aunio and Hautamäki (2005) applying only Let's Think! (Adey et al., 2001) showed no intervention effects in the immediate post-test, it is at least worth suggesting that the design combining the two programmes may not have been so unsuccessful after all.

Cognitive factors. The research scope could have been broadened by including two concepts in early-childhood mathematical development, namely subitising ability and spontaneous focusing on numbers (Hannula, 2005). Subitising (e.g., Wynn, 1998) is the ability to enumerate small sets of one to five objects without counting them, while spontaneous focusing on numbers refers to children focusing their attention on numbers of objects or incidents, and numerosity is considered a relevant factor in tasks in which the numeric aspect is immediately obvious to adults (Hannula, 2005). Both of these skills play a role in the development of number sense as measured in this study, and it is recommended that these elements should be integrated in future studies.

There are also other important cognitive elements of number sense that were not investigated here. It would be useful to know to what extent language skills explain children's number-sense test performance, when there is a need to measure the early numerical skills of children with language difficulties, or whose native language is other than Finnish. The complexity of number sense in terms of the cognitive factors related to it has also been demonstrated (Kyttälä, Aunio, Lehto, Van Luit, & Hautamäki, 2003), as it was found that preschooler's relational skills correlated with general intelligence, and counting skills with visuospatial working-memory capacity. Given the limitations of that study, there will be a new research to examine further the effects of the underlying cognitive system on young children's number-sense development. It will investigate the language and visuospatial systems that are responsible for children's abilities to represent and manipulate numerical information (Geary, 2004). This kind of approach will enhance understanding of the difficulties that children encounter in their early number-sense development.

Cultural factors. One element that is likely to affect early number-sense development is the home learning environment (Aubrey, Bottle, & Godfrey, 2003), and parental support of learning mathematics in particular. This study was restricted to some speculative statements on the issue: given the cultural background, for example, it seemed rational to suggest that parents in Asian locations appreciate mathematical skills, know how, and are willing to support such learning at home. In order to find out more, it would be useful to engage in subsequent cross-national comparison research into young children's number sense, which would incorporate the possibility of controlling the effects of parental educational support. It would also be interesting to make cross-national comparisons inside Europe, in which Finnish children would also be included (see Torbeyns et al., 2002; Van de Rijt et al., 2003 for some European comparisons without Finnish children). Comparison with children from the United Kingdom and Finland would also be an interesting avenue, since as the children in the United Kingdom begin their compulsory education at the age of five whereas in Finland children start primary school at the age of seven. It would be even more appealing to include data about the home-learning environment and children's development in longitudinal studies.

Motivational factors. Individual motivation plays a vital role in learning and achievement behaviour, such as in choosing tasks, paying attention, expending effort, and showing persistence (Spinath, 2005), and a lack of task motivation in mathematics has been found to contribute to slow development during the first three years of primary school (Nurmi & Aunola, 2005). Motivational factors play an important role in terms of the effect of interventions on mathematical thinking. The way in which children participate in the sessions is dependent on how motivated they are. The motivation to learn is often particularly low in low-performance children, and this is thus clearly a factor that needs to be included in future interventions.

In brief, the future research challenges lie in explaining the number-sense differences in young children in terms of cognitive, attentional, cultural and motivational factors, which in turn will provide the necessary information to enable educational-support decisions to be made.

• • •

In conclusion, this study benefits the scientific field of early-childhood (special) education and mathematical-knowledge development in at least five ways. First, it demonstrates the applicability of the Finnish Early Numeracy Test (Van Luit et al., 2005), which can now be used by educational professionals to measure the mathematical skills of Finnish children aged between four and seven years. It also shows that it is possible to support young children's mathematical development by applying carefully planned instructional tools. Thirdly, it provides evidence that children with a multilanguage background and SEN children already lag behind their peers in number-sense development in preschool years. Fourthly, the study demonstrated and proved the validity of a new internationally relevant research application of the Early Numeracy Test, in that by using two separate scales measuring children's relational and counting skills, it was possible to collect different information than in analysis relying on one total score. Finally, this study broadened our knowledge about international differences in children's early-mathematical knowledge since no previous studies had included Finnish children as participants.

Notes

1 Reprints were made with the kind permission of the publishers.
2 In this doctoral dissertation the term preschool education refers to the early childhood education of children from one to six years old.
3 The counting scale was called the number-sequence scale in the original publication.
4 Young Children with Special Educational Needs Count Too! (Count too!; Van Luit & Schopman, 1998) was referred to as Math! in the original publication.

References

Adey, P. (1997). CASE, Cognitive acceleration through science in education. In J. H. M. Hamers & M.Th. Overtoom (Eds.), *Teaching thinking in Europe: Inventory of European programmes* (pp. 175–178). Utrecht, The Netherlands: Sardes.

Adey, P., Robertson, A., & Venville, G. (2001). *Let's think! A programme for developing thinking in five and six year olds*. Windsor, UK: Nfer-Nelson.

Adey, P., Robertson, A., & Venville, G. (2002). Effects of a cognitive acceleration programme on year 1 pupils. *British Journal of Educational Psychology, 72*, 1–25.

Aguilar, M., Navarro, J. I., Alcalde, C., Ruiz, G., & Marchena, E. (2005, August). *The Utrecht Early Mathematical Competence test in Spanish sample.* Poster presented at the 11th conference of European Association for Research on Learning and Instruction (EARLI). August 23-27. Nicosia, Cyprus.

American Educational Research Association, American Psychological Association, National Council on Measurement in Education (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

Arvedson, P. J. (2002). Young children with specific language impairment and their numerical cognition. *Journal of Speech, Language, and Hearing Research, 45,* 970–982.

Aubrey, C. (1999). *Developmental approach to early numeracy*. Birmingham, UK: Questions.

Aubrey, C. (2001). Early mathematics. In T. Davia (Ed.), *Promoting evidence based practice in early childhood education* (pp. 185–210). London, UK: Elsevier Science.

Aubrey, C., Bottle, J. & Godfrey, R. (2003). Early mathematics in home and out-of-home context. *International Journal of Early Years Education 11,* 91–103.

Aunio, P., & Hautamäki, J. (2005, August). *The effects of Let's Think-programme in early mathematical thinking abilities—results from three Finnish studies*. Paper presented at the 11th conference of European Association for Research on Learning and Instruction (EARLI). August 23-27. Nicosia, Cyprus.

Baker, F. (2001). *The basics of item response theory*. College Park, MD: ERIC Clearing House on Assessment and Evaluation. University of Maryland.

Bryant, P. (1996). Children and arithmetic. In L. Smith (Ed.), *Critical readings on Piaget* (pp. 312–346). London, UK: Routledge.

Bryant, P., & Nunes, T. (2002). Children's understanding of mathematics. In U. Goswami (Ed.), *Blackwell handbook of childhood cognitive development* (pp. 412–439). Malden, The Netherlands: Blackwell.

Butterworth, B. (1999). *The mathematical brain.* Oxford, UK: Macmillan.

Campbell, J., & Xue, Q. (2001). Cognitive arithmetic across cultures. *Journal of Experimental Psychology: General, 130*, 299–315.

Caplan, N., Choy, M., & Whitmore, J. (1992). Indochinese refugee families and academic achievement. *Scientific American, 266*, 36–42.

Carr, M., & Jessup, D. (1997). Gender differences in first grade mathematics strategy use: Social and metacognitive influences. *Journal of Educational Psychology, 89*, 318–328.

Case, R. (1996). Reconceptualizing the nature of children's conceptual structures and their development in middle childhood. In R. Case & Y. Okamoto (Eds.), The role of central conceptual structures in the development of children's thought, *Monographs of the Society for Research in Child Development, 61* (1-2), Serial No. 246, 1–26.

Case, R., & Okamoto, Y. (Eds.) (1996). The role of central conceptual structures in the development of children's thought, *Monographs of the Society for Research in Child Development, 61* (1-2), Serial No.246.

City of Helsinki, (2004). *Central features of social services in the operational year 2004*. Helsinki, Finland: Social Service Department.

Cooper, H., Charlton, K., Valentine, J. C., & Muhlenbruck, L. (2000). Making the most of summer school: a meta-analytic and narrative review. *Monographs of the Society for Research in Child Development, 65* (1), Serial No. 260.

Dehaene, S. (1997). *The number sense. How the mind creates mathematics*. London, UK: Penguin Books.

DeShazo-Barry, T., Lyman, R. D., & Grofer-Klinger, L. (2002). Academic underachievement and attention-deficit/hyperactivity disorder: The negative impact of symptom severity on school performance. *Journal of School Psychology, 40*, 259–283.

Demie, F. (2001). Ethnic and gender differences in educational achievement and implications for school improvement strategies. *Educational Research, 43* (1), 91–106.

Donlan, C. (Eds) (1998). *The development of mathematical skills*. Hove, UK: Psychology Press.

Fazio, B. B. (1999). Arithmetic calculation, short-term memory and language performance in children with specific language impairment: A 5-year follow-up. *Journal of Speech, Language and Hearing Research, 42,* 420–421.

Fuson, K. (1988). *Children's counting and concept of number*. New York: Springer Verlag.

Fuson, K., & Kwon, Y. (1992). Learning addition and subtraction: Effects of number words and other cultural tools. In J. Bideaud, C. Meljac, & J-P. Fisher (Eds.), *Pathways to number: Children's developing numerical abilities* (pp. 283–306). Hillsdale, NJ: Erlbaum.

Fuson, K. C., Carrol, W. M., & Drueck, J. V. (2000). Achievement results for second and third graders using the Standards-based curriculum Every Day Mathematics. *Journal for Reacearch in Mathematics Education, 31* (3), 277–295.

Fuson, K. C, Smith, S. T, & Lo Cicero, A. M. (1997). Supporting Latino first graders' ten-structured thinking in urban classrooms. *Journal for Research in Mathematics Education, 28* (6), 738–760.

Galperin, P. (1969). Stages in the development of mental acts. In Cole & Meltzman (Eds.), *Handbook of contemporary Soviet psychology* (pp. 249–273). New York, Basic Books.

Geary, D. C. (1994). *Children's mathematical development*. Washington, DC:APA.

Geary, D. C. (2000). From infancy to adulthood: the development of numerical abilities. *European Child & Adolescent Psychiatry, 9*: II/11-II/16.

Geary. D. C. (2004). Mathematics and learning difficulties. *Journal of Learning Disabilities, 37* (1), 4–15.

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

Ginsburg, H., Choi, Y., Lopez, L., Netly, R., & Chi, C. (1997). Happy birthday to you: Early mathematical thinking of Asian, South American and U.S. children. In T. Nunes & P. Bryant (Eds.), *Learning and teaching mathematics: An international perspective* (pp. 163–207). East Sussex, UK: Erlbaum Taylor & Francis.

Gorard, S., Rees, G., & Salisbury, J. (2001). Investigating the patterns of differential attainment of boys and girls at school. *British Educational Research Journal, 27* (2) 125–139.

Griffin, S. (2000). *Number worlds: Preschool level*. Durham, NH: Number Worlds Alliance.

Griffin, S. (2003). The development of math competence in the preschool and early school years. In J. M. Royer (Ed.) *Mathematical cognition* (pp. 1–32). Greenwich, CN: Information age publishing.

Griffin, S. & Case, R. (1995). *Number worlds: Kindergarten level*. Durham, NH: Number Worlds Alliance.

Griffin, S., & Case, R. (1998). Re-thinking the primary school math curriculum: An approach based on cognitive science. *Issues in Education, 4,* 1–51.

Hamers, J. H. M., & Overtoom M. Th., (Eds.) (1997). *Teaching thinking in Europe: Inventory of European programmes.* Utrecht, The Netherlands: Sardes.

Hannula, M. M., (2005). *Spontaneous focusing on numerosity in the development of early mathematical skills.* Doctoral dissertation. Annales Universitatis Turkuensis, Serie B 282. Turku, Finland: University of Turku.

Hannula, M. M., & Lehtinen, E. (2005). Spontaneous focusing on numerosity and mathematical skills of young children. *Learning and Instruction, 15*, 237–256.

Hautamäki, J. (1984). *Peruskoululaisten loogisen ajattelun mittaamisesta ja esiintymisestä* [The measurement and distrubution of Piagetian stages of thinking in Finnish comprehensive school]. Doctoral dissertation. Publications in Social Sciences 1/1984. Joensuu, Finland: University of Joensuu.

Hautamäki, J., Bergholm, B., Hautamäki, A., Lehto, J., Kupiainen, S., Outinen, K., & Pekuri, M. (2001). *Ensiaskeleet – oppimisen edellytykset. Luokanopettajille tarkoitetun seulan toimivuus Helsinki-aineiston perusteella* [Preacademical abilities – a scale for primary educators. Norms from Helsinki]. Research Reports A17:2001. Helsinki, Finland: Helsinki City Board of Education.

Helland, T., & Asbjornsen, A. (2003). Sequential and visuo-spatial skills in dyslexia: variations according to language comprehension and mathematical skills. *Child Neuropsychology, 9,* 208–220.

Hosenfeld, B., Van den Boom, D. C., & Resing, W. C. M. (1997). Constructing geometric analogies for the longitudinal testing of elementary school children. *Journal of Educational Measurement, 34* (4), 367–372.

Huntsinger, C., Jose, P., Liaw, F-R., & Ching, W-D. (1997). Cultural differences in early mathematics learning: A comparison of Euro-American, Chinese-American, and Taiwan-Chinese families. *International Journal of Behavioral Development, 21,* 371–388.

Ikäheimo, H. (1996). *Matematiikan keskeisten käsitteiden diagnoosi esiopetuksen alussa ja lopussa sekä 1.luokan alussa* [The assessment of mathematical basic concepts in the beginning of the preschool and primary grade]. Helsinki, Finland: Opperi.

Jose, P., Huntsinger, C., Huntsinger, P., & Liaw, F-R. (2000). Parental values and practices relevant to young children's social development in Taiwan and the United States. *Journal of Cross-Cultural Psychology, 31,* 677–702.

Kaminen, L. (2004). Kuntoutusväline lukukäsitteen kehittämiseen – Tapaustutkimus Jannen matematiikan oppimisvaikeuksista, lukutaidon kehittymisestä ja tarkkaavaisuudesta. [Intervention method for supporting number sense development—case study in learning difficulties, number sense development and attention.] Unpublished bachelor thesis. Helsinki, Finland: University of Helsinki.

Kautonen, P. (1999). *Lasten matemaattiset taidot esikoulussa ja ensimmäisellä luokalla. Seuranta Utrechtin lukukäsitetestillä.* [Children's mathematical abilities in preschool and primary one: A follow-up study with Early Numeracy Test]. Unpublished master thesis. Joensuu, Finland: University of Joensuu.

Keranto, T. (1981). Lukukäsitteen kehittyminen ja kehittäminen: Matemaattis-loogiset perusteet ja luvun kognitiivinen rakentuminen [The aquisition and aided development of the concept of number: mathematical-logical principles and the cognitive structuring of number]. *Acta Universitatis Tamperensis* A 125. Tampere, Finland: University of Tampere.

Kinnunen R., Lehtinen, E., & Vauras, M. (1994). Matemaattisen taidon arviointi [Assessment of mathematical skills]. In M. Vauras, E. Poskiparta, & P. Niemi (Eds.*), Kognitiivisten taitojen ja motivaation arviointi koulutulokkailla ja 1. luokan oppilailla* [Assessment of coginitive skills and motivation in school beginners and first graders] (pp. 55–76). Turku, Finland: Univerisity of Turku, Centre for Learning Research.

Koponen, T., Mononen, R., & Räsänen, P. (2003). Tutkimus 9–11 vuotiaiden dysfaattisten lasten peruslaskutaidoista [A research on basic arithmetic skills of 9–11 years-old dyslexia children]. *NMI-bulletin,* 13 (4), 3–8.

Korkman, M., Kirk, U., & Kemp, S. L. (1997). *NEPSY – Lasten neuropsykologinen tutkimus* [Neurocognitive measures for children]. Helsinki, Finland: Psykologien Kustannus Oy.

Kroesbergen, E. H., & Van Luit, J. E. H. (2003). Mathematics interventions for children with special educational needs: A meta-analysis. *Remedial & Special Education, 24*, 97–114.

Kuusela, J. (2000). *Tieteellisen paradigman mukaisen ajattelun kehittyminen peruskoulussa – kahden interventiomenetelmän vertaileva tutkimus peruskoulun kuudesluokkalaisilla* [The development of thinking towards a scientific paradigm in the comprehensive school—A comparative study of two intervention programmes for sixth grade students]. Doctoral dissertation. Department of Teacher Education, Research Report 221. University of Helsinki, Finland.

Kyttälä, M., Aunio, P., Lehto, J. E., Van Luit, J. E. H., & Hautamäki, J. (2003). Visuospatial working memory and early numeracy. *Educational and Child Psychology, 20* (3) 65–76.

Li, J. (2002). A cultural model of learning: Chinese "hart and mind wanting to learn". *Journal of Cross-Cultural Psychology, 33,* 248–269.

Li, J. (2004). A Chinese cultural model of learnig. In L. Fan, N-Y. Wong, J. Cai, & S. Li (Eds.), *How Chinese learn mathematics—perspective from insiders* (pp. 124–156). Series of Mathematics Education Vol 1. Singapore: World Scientific Publishing.

Marshall, R. M., Hynd, G. W., Handwerk, M. J., & Hall, J. (1997). Academic underachievement in ADHD subtypes. *Journal of Learning Disabilities, 30*, 635–642.

Miller, K., Smith, C., Zhu, J., & Zhang, H. (1995). Preschool origins of cross-national differences in mathematical competence: The role of number-naming system. *Psychological Science, 6,* 56–60.

Miura I. (1987). Mathematics achievement as a function of language. *Journal of Educational Psychology. 79*, 79–82.

Miura, I., Okamoto, Y., Kim, C., Steere, M., & Fayol, M. (1993). First graders' cognitive representation of number and understanding of place value: Cross-national comparisons—France, Japan, Korea, Sweden and the United States. *Journal of Educational Psychology. 85,* 24–30.

Miura, I., Okamoto, Y., Vlahovic-Stetic, V., Kim, C., & Han, J. (1999). Language supports for children's understanding of numerical fractions: Cross-national comparisons. *Journal of Experimental Child Psychology [on-line] 74*. Available in http://www.idealibrary.com, Item: 1999.2519.

Miura, I., Okamoto, Y., Kim, C., Vlahovic-Stetic, V., Han, J., Lemaire, P., & Hatano, E. (2000, July). *The development of numerical fraction understanding: Cross-national comparison*. Poster session presented at the 16th Biennial Meeting of ISSBD, Beijing, PRC.

Mullins, I. V. S., Martin, M. O., Gonzalez, E. J., & Chorostowski, S. J. (2004). *Findings from IEA's Trends in International mathematics and science study at the fourth and eight grades.* Chestnut Hill, MD: TIMSS & PIRLS international study center, Boston College.

Muthén, B. (1989). Latent variable modeling in heterogeneous populations. *Psychometrika, 54*(4), 557–585.

Muthén, L. K. & Muthén, B.O. (1998–2004). *Mplus user's guide, Third edition*. Los Angeles, CA: Muthén & Muthén.

Nunes T. & Bryant P. (1996). *Children doing mathematics.* Oxford, UK: Blackwell.

Nurmi, J.-E., & Aunola, K. (2005). Task-motivation during the first school years: A person-oriented approach to longitudinal data. *Learning and Instruction, 15,* 103–122.

Okamoto, Y., Miura, I., Suomala, J., & Curtis, R. (1996, April). *Developmental differences in children's cognitive representation of number and their understanding of place value: A cross-national comparison.* Paper presented at the annual meeting of the American Educational Research Association, New York.

Perry, M., VanderStoep, S., & Yu, S. (1993). Asking questions in first-grade mathematics classes potential influence on mathematical thought. *Journal of Educational Psychology, 85*, 31–40.

Piaget, J. (1965). *The child's conception of number*. New York: Norton.

Piaget, J. & Inhelder, B. (1974). *The child's construction of quantities.* London, UK: Routledge & Kegan Paul.

Piaget, J. & Inhelder, B. (1976). *The child's conception of space*. London, UK: Routledge & Kegan Paul.

PISA (2003). *Learning for tomorrows world—First results from PISA 2003.* OECD Publishing.

Rittle-Johnson, B. & Siegler, R. (1998). The relation between conceptual and procedural knowledge in learning mathematics. In Donlan, C. (Eds), *The development of mathematical skills* (pp. 75–109). Hove, UK: Psychology Press.

Ruijssenaars, A.J.J.M., Van Luit, J.E.H., & Van Lieshout, E.C.D.M. (2004). *Rekenproblemen en dyscalculie – theorie, onderzoek, diagnostiek en behandeling.* Rotterdam, The Netherlands: Lemniscaat.

Räsänen, P. (2005). *BANUCA* [Basic numerical and calculation abilities]. Jyväkylä, Finland: Niilo Mäki Instituutti.

Sammons, P. & Smees, R. (1998). Measuring pupil progress at Key Stage 1: using baseline assessment to investigate valued added. *School Leadership & Management, 18* (3), 389–407.

Schopman, E.A .M., & Van Luit, J.E.H. (1995). Young children's counting strategies: an observational study. *European Early Childhood Education Research Journal, 3* (2), 67–78.

Schopman, E.A .M., Van Luit, J.E.H. & Van de Rijt, B.A.M. (1996). Learning and transfer of preparatory arithmetic strategies among young children with developmental lag. *Journal of Cognitive Education, 5,* 117–131.

Secada, W. G. (1991). Degree of bilingualism and arithmetic problem solving in Hispanic first graders. *The Elementary School Journal, 92* (2), 213–231.

Sharpe, P. (2002). Preparing for primary school in Singapore—Aspects of adjustment to the more formal demands of the primary one mathematics syllabus. *Early Childhood Development and Care, 172* (4) 329–335.

Shayer, M. & Adhami, M. (2003). Realising the cognitive potential of children 5–7 with a mathematics focus. *International Journal of Educational Research, 39*, 743–775.

Shayer, M. & Wylam, H. (1978). The distribution of Piagetian stages of thinking in British middle and secondary school children. II: 14–16 year-olds and sex differentials. *British Journal of Educational Psychology, 48,* 62–70.

Smith, L. (2002). *Reasoning by mathematical induction in children's arithmetic.* Oxford, UK: Pergamon Press.

Song, M-J., & Ginsburg, H. (1987). The development of informal and formal mathematical thinking in Korean and U.S. children. *Child Development, 58,* 1286–1296.

Sophian, C. (1998). A developmental perspective on children's counting. In C. Donlan (Ed.), *The development of mathematical skills* (pp. 27–46). Hove, UK: Psychology Press.

Spinath, B. (2005). Development and modification of motivation and self-regulation in school contexts: Introduction to the special issue. *Learning and Instruction, 15*, 85–86.

Stevenson, H. & Lee, S-Y. (1995). The East Asian version of whole-class teaching. *Educational Policy, 9,* 152–168.

Stevenson, H., Lee, S., & Graham, T. (1993). Chinese and Japanese kindergartens: Case study in comparative research. In B. Spodek (Ed.), *Handbook of research on the education of young children* (pp. 519–535). New York: Macmillan.

Stigler, J. & Perry, M. (1990). Mathematics learning in Japanese, Chinese and American classrooms. In J. Stigler, R. Shweder, & G. Herdt (Eds.), *Cultural psychology: Essays on comparative human development* (pp. 328–353).New York, Cambridge University Press.

Strand, S. (1997). Pupil progress during Key Stage 1: a value-added analysis of school effects. *British Educational Research Journal, 23* (4), 471–487.

Strand, S. (1999). Ethnic group, sex, and economic disadvantage: associations with pupils' educational progress from Baseline to the end of Key Stage 1. *British Educational Research Journal, 25* (2), 179–202.

Tabachnick, B. G., & Fidell, L. S. (1996) *Using multivariate* statistics, 3th ed. New York: Harper & Row.

Torbeyns, J. (1999). *Vlaamse normering van de Utrechtse Getalbegrip Toets* [The construction of the norms for the Early Numeracy Test for Flanders; in Dutch]. Unpublished master's thesis. Leuven, Belgium: University of Leuven.

Torbeyns, J., Van de Noortgate, W., Ghesquière, P., Verschaffel, L. Van de Rijt, B. A. M., & Van Luit, J. E. H. (2002). The development of early mathematical competence of 5- to 7-year-old children—A comparison between Flanders and The Netherlands. *Educational Research and Evaluation, 8,* 249–275.

Tuss, P., & Zimmer, J. (1995). Causal attributions of underachieving fourth-grade students in China, Japan and the United States. *Journal of Cross-Cultural Psychology, 26,* 408–426.

Tzouriadou, M., Barbas, G., & Bonti, E. (2002). Socio-cultural environment as a factor of differentation in mathematical reasoning. *Psychology, 9,* 281–294.

Van de Rijt, B. A. M. (1996*). Voorbereidende rekenvaardigheiden bij kleuters* [Early mathematical competence in young children]. Doctoral dissertation. Utrecht University. Doetinchem, The Netherlands: Graviant.

Van de Rijt, B. A. M , Godfrey, R., Aubrey, C., Van Luit, J. E. H., Ghesquiere, P., Torbeyns, J. L., Hasemann, K., Tancig, S., Kvakler, M., Magajna, L., & Tzouriadou, M. (2003). The development of early numeracy in Europe. *Journal of Early Childhood Research, 1,* 155–180.

Van de Rijt, B. A. M. & Van Luit, J. E. H. (1993, August). *The results of different treatments on children's weak performance in preparatory and initial arithmetic*. Paper presented at the 5th European Conferencce for Research on Learning and Instruction. August 31-September 5, Aix-en-Provence, France.

Van de Rijt, B. A. M. & Van Luit J. E. H. (1998). Effectiveness of the additional early mathematics program for teaching children early mathematics. *Instructional Science 26*, 337–358.

Van de Rijt, B. A. M. & Van Luit, J. E. H. (1999). Milestones in the development of infant numeracy. *Scandinavian Journal of Psychology, 40*, 65–71.

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

Van de Vijver, F. & Leugn, K. (2000). Methodological issues in psychological research on culture. *Journal of Cross-Cultural Psychology, 31*, 33–51.

Van Luit, J. E. H, & Schopman, E. (1998). *Als speciale kleuter tel je ook mee! Voorbereidende rekenacriviteiten voor kleuters met een ontwikkelingsachterstand* [Young children with special educational needs count, too! Early mathematics activities for young children with special educational needs]. Utrecht, The Netherlands: Graviant.

Van Luit, J. E. H., & Schopman, E. A. M. (2000). Improving early numeracy of young children with special educational needs. *Remedial & Special Education, 21* (1), 27–40.

Van Luit, J. E. H., Van de Rijt, B. A. M., & Aunio, P. (2003/5) *Early Numeracy Test, Finnish Edition* [Lukukäsitetesti]. Helsinki, Finland: Psykologien kustannus.

Van Luit, J. E. H., Van de Rijt, B. A. M., & Hasemann, K. (2001). *OTZ Osnabrücker Test zur Zahlbegriffsentwicklung* [German version of the Utrecht Test of Number Sense]. Göttingen, Germany: Hogrefe-Verlag.

Van Luit, J. E. H., Van de Rijt, B. A. M., & Pennings, A. H. (1994). *Utrechtse Gatalbegrip Toets, UGT* [Early Numeracy Test]. Doetinchem, The Netherlands: Graviant.

Vauras, M., Poskiparta, E., & Niemi, P. (Eds.) (1994). *Kognitiivisten taitojen ja motivaation arviointi koulutulokkailla ja 1. luokan oppilailla* [The assessment of cognitive skills and motivation of school beginners and first graders]. Turku, Finland: University of Turku, Centre for Learning Research.

Verschaffel, L. (2005, August). *Conceptual understanding and procedural skills in mathematics learning: Discussion.* Discussion paper presented at the 11th conference of the European Association for Research on Learning and Instruction. August 23-27. Nicosia, Cyprus.

Vygotsky, L.S. (1978). *Mind in society.* Cambridge, MIT Press.

Wechsler, D. (1995). *Wechsler Preschool and Primary School Scale of Intelligence*. Finnish Edition. Helsinki, Finland: Psykologien Kustannus Oy.

Wertheimer, M. (1971). *Productive thinking*. New York, Harper & Row.

Wynn, K. (1998). Numerical competence in infants. In C. Donlan (Ed.), *The development of mathematical competence* (pp. 3–25). Hove, UK: Psychology Press.

Zhou, Z., & Boehm, A. (2001). American and Chinese children's knowledge of basic relational concepts. *School Psychology International, 22*, 5–21.

Zhou, Z., Cheng, C., Mottram, L., & Rosenblum, S. (1999, April). *Understanding informal and formal mathematical abilities in Mainland Chinese and Chinese-American children.* Paper presented at the Biennial Meeting of the Society for Research in Child Development, Albuquerque, US.

Yang, H. & al. (1985). The development of number concept on young children. *Information on Psychological Sciences, 5,* 57–59. [The articles name is translated from Chinese]

Yoshida, M., Fernandez, C., & Stigler, J. (1993). Japanese and American students' differential recognition memory for teachers' statements during a mathematics lesson. *Journal of Educational Psychology, 85*, 610–617.