METADATA OF THE STUDY:

Interaction, Development & Learning (VUOKKO)

A Longitudinal Study from Toddlerhood to Grade 3

15.5.2024

Tiia Nurminen, Minna Torppa, Tuire Koponen, Kenneth Eklund & Jenni Salminen

University of Jyväskylä

JYU.
EDU

About this metadata

VUOKKO researchers are aiming toward open collaboration with the research field and welcome individual researchers to jointly examine the data and seek opportunities for collaboration. The collaboration starts by discussions with the steering group of the VUOKKO dataset and by signing the VUOKKO authorship agreement. The agreement explicates the code of conduct for managing the longitudinal dataset, distributing access to datasets, agreeing on authorship roles and informing the research group members about research ideas.

Contact information:

- For more information, please, contact the VUOKKO researchers through email: vuokko-tutkimus@jyu.fi
- Link to the VUOKKO website: www.jyu.fi/vuokko
- Link to the VUOKKO in Converis: https://converis.jyu.fi/converis/portal/detail/ResearchDataset/177146477

How to cite VUOKKO study (APA 7):

- Use this citation to refer to this document:
  - Nurminen, T., Torppa, M., Koponen, T., Eklund, K., & Salminen, J. (2024). *Metadata of the Study: Interaction, Development & Learning (VUOKKO) - A Longitudinal Study from Toddlerhood to Grade 3*. University of Jyväskylä. https://doi.org/10.17011/jyx/dataset/94856
- Use this citation for the **ECEC phase** of VUOKKO (referring to the data used):
  - Lerkkanen, M.-K., & Salminen, J. (2015–2019). *Interaction, Development, and Learning (VUOKKO) -study: Early Childhood Education* [Research data]. University of Jyväskylä. Finland.
- Use this citation for the **school phase (Grades 1–3)** of VUOKKO (referring to the data used):
  - Salminen, J., Lerkkanen, M.-K., Koponen, T., & Torppa, M. (2021–2023). *Interaction, Development, and Learning (VUOKKO) -study: Early School Years (Grades 1–3)* [Research data]. University of Jyväskylä. Finland.
- Use this citation for the **school phase (Grades 4–5)** of VUOKKO (referring to the data used):
  - Salminen, J., Ruotsalainen, J., Koponen, T., Lerkkanen, M.-K., & Torppa, M. (2023–2025). *Interaction, Development, and Learning (VUOKKO) -study: Early School Years (Grades 4–5)* [Research data]. University of Jyväskylä. Finland

The metadata file was written with the support of EarlyMath-project, which has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 101002966).

Contents

1. Description of the Study 1
   1.1. Funding 3
   1.2. Publications 3
   1.3. Creators 3
   1.4. Access to Data 5
2. Ethical Considerations 5
3. Data Management Principles 8
4. Data Collection 10
   4.1. RP – Recruitment and Participants 10
      4.1.1. RP – Toddler Phase (2015–2016, T1–T2) 10
      4.1.2. RP – Preschool Phase (2018–2019, T3–T4) 12
      4.1.3. RP – School Phase: Grade 1 (2020–2021, T5) 14
      4.1.4. RP – School Phase: Grade 2 (2021–2022, T6) 15
      4.1.5. RP – School Phase: Grade 3 (2022–2023, T7) 17
      4.1.6. RP – Parental Assessments (2021–2022, T5–T6) 19
      4.1.7. Sample Size at Each Timepoint and Attrition 20
   4.2. Procedures 22
      4.2.1. CA – Children's Assessments 22
      4.2.2. CR – Children's Reports 29
      4.2.3. PA – Parental Assessments 30
      4.2.4. PQ – Parental Questionnaires 31
      4.2.5. TCI – Teacher–Child Interaction Quality (Classroom Observations) 33
      4.2.6. TQ – Teachers' Questionnaires 39
      4.2.7. TR – Teachers' Reports Concerning Individual Children 40
5. Instruments in Data Collection / Measures 41
   5.1. CA – Children's Assessments 41
      5.1.1. CA – Toddler Phase (T1–T2) 44
      5.1.2. CA – Preschool Phase (T3–T4) 49
      5.1.3. CA – School Phase: Grade 1 (T5) 55
      5.1.4. CA – School Phase: Grade 2 (T6) 59
      5.1.5. CA – School Phase: Grade 3, Group Assessment (T7) 66
      5.1.6. CA – School Phase: Grade 3, Individual Assessment (T7) 72
   5.2. CR – Children's Reports 75
      5.2.1. CR – Preschool Phase (T3–T4) 75
      5.2.2. CR – School Phase: Grade 2 (T6) 76

5.2.3. CR – School Phase: Grade 3 (T7) 77
5.3. RR – Research Assistants' Reports Concerning Individual Child 79
5.3.1. RR – Preschool Phase (T3–T4) 79
5.3.2. RR – School Phase: Grade 3, Individual Assessment (T7) 79
5.4. PA – Parental Assessments 80
5.5. PQ – Parental Questionnaires 86
5.5.1. PQ – Toddler Phase (T1) 89
5.5.2. PQ – Preschool Phase (T3) 92
5.5.3. PQ – School Phase: Grade 1 (T5) 97
5.5.4. PQ – School Phase: Grade 2 (T6) 106
5.5.5. PQ – School Phase: Grade 3 (T7) 115
5.6. TCI – Teacher–Child Interaction Quality (Classroom Observations) 125
5.6.1. TCI – Toddler Phase (T1–T2) 125
5.6.2. TCI–AC – Toddler Phase (T1–T2) 126
5.6.3. TCI – Preschool Phase (T3–T4) 127
5.6.4. TCI – School Phase: Grade 1 (T5) 128
5.6.5. TCI – School Phase: Grade 2 (T6) 129
5.6.6. TCI – School Phase: Grade 3 (T7) 129
5.7. TQ – Teachers' Questionnaires 130
5.7.1. TQ – Toddler Phase (T1–T2) 133
5.7.2. TQ – Preschool Phase (T3–T4) 138
5.7.3. TQ – School Phase: Grade 1 (T5) 146
5.7.4. TQ – School Phase: Grade 2 (T6) 149
5.7.5. TQ – School Phase: Grade 3 (T7) 153
5.8. TR – Teachers' Reports Concerning Individual Children 156
5.8.1. TR – Toddler Phase (T1–T2) 157
5.8.2. TR – Preschool Phase (T3–T4) 158
5.8.3. TR – School Phase: Grade 1 (T5) 160
5.8.4. TR – School Phase: Grade 2 (T6) 161
5.8.5. TR – School Phase: Grade 3 (T7) 162
6. Coding Protocols 165
7. Description of Variables in the Data Files 166
7.1. ID Creation Protocol 166
7.2. Variable Naming Protocol 166
7.3. Variable Labelling Protocol 167

1. Description of the Study

VUOKKO is an early onset longitudinal study from age 2 to age 9. The main aim of the VUOKKO study is to better understand the developmental paths of children's skills, motivation, and emotions in the domains of language, reading, numeracy, and math as well as the interactions in the learning environments. The early onset of cognitive skills along with social and self-regulation skills are examined in addition to indicators of learning motivation and emotions. The study also aims to gain more specified understanding on the role that the home environment and parental skills play in children's development early in childhood and later in school. Finally, the study aims to examine associations between the quality of teacher–child interactions and child outcomes (early academic, social and self-regulation skills) both concurrently and across time.

This document describes the data of the VUOKKO study from toddlerhood to Grade 3 (T1–T7). The longitudinal data include data related to children, parents and teachers. Figure 1 gives an overview of the research design. The data include the following:

- Data collected from **the children[1]:**
  - Children's numeracy and literacy skills
  - Children's reports
- Data collected from **the parents**[2]:
  - Parental math and literacy skills
  - Parental questionnaires concerning both the child and parent
- Data collected from **the teachers**[3]:
  - Teacher–child interaction in ECEC classrooms and in primary school classrooms
  - Teachers' questionnaires
  - Teachers' reports concerning the individual child

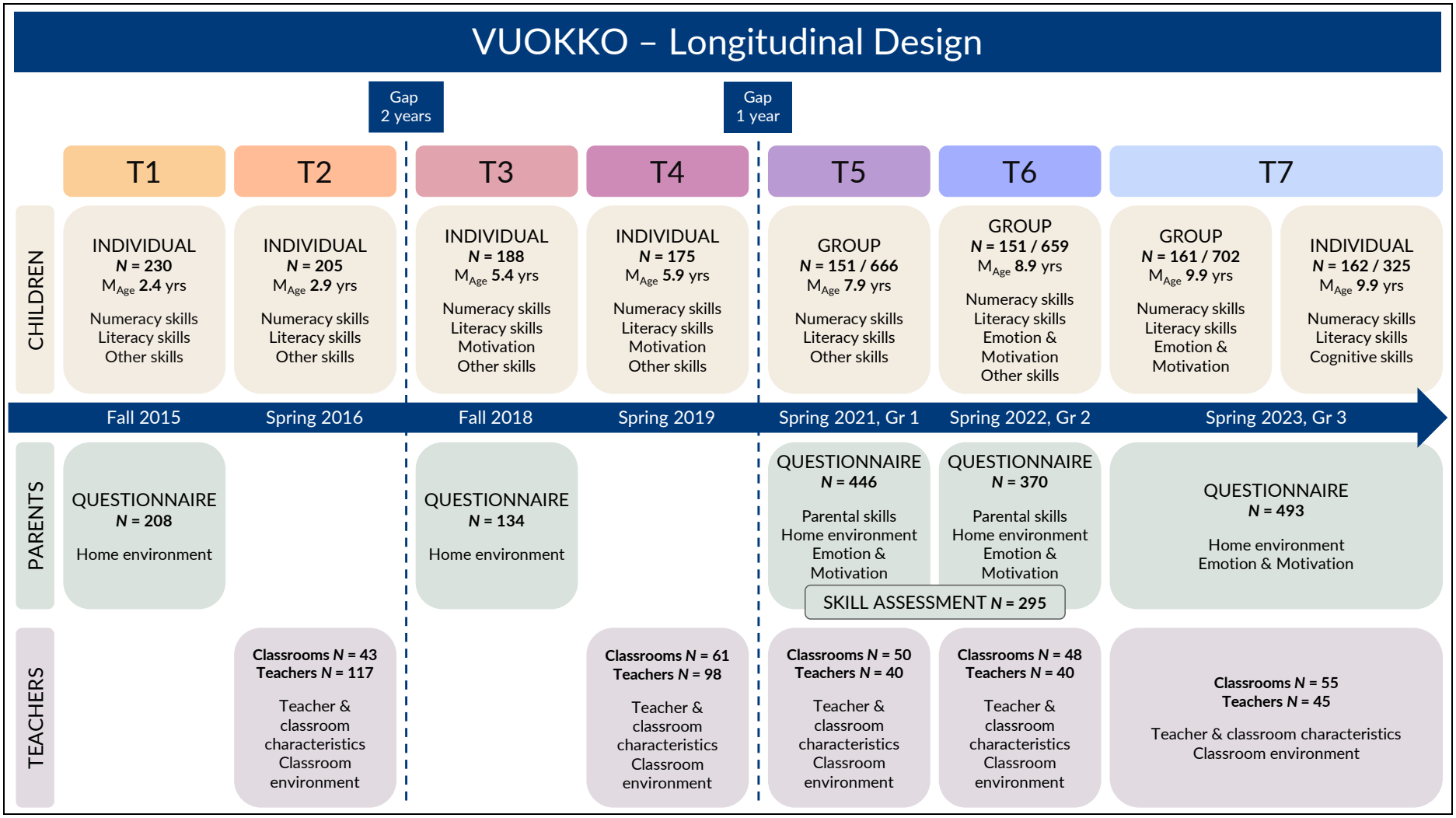
Data gathering will continue in 2024–2025 in Grades 4 and 5 (T8–T9). Metadata of those timepoints will be published later.

[1] Data is collected from the *children* in seven timepoints in the following ages. T1: 2.4 years, T2: 2.9 years, T3: 5.4 years, T4: 5.9 years, T5: 7.9 years (Grade 1), T6: 8.9 years (Grade 2), and T7: 9.9 years (Grade 3).

[2] *Parent* is used as a generic term across this metadata file to refer either to a child's biological parent or to another non-biological parent or guardian.

[3] *Teacher* is used as a generic term across VUOKKO samples to refer to the variety of professional educators working with children both in ECEC settings and later in schools. The term *teacher* is therefore also used when referring to ECEC staff members in ECEC classrooms. The staff comprises teachers in ECEC, social pedagogues in ECEC, childcarers in ECEC and the special education teachers in ECEC all engaged to work in the participating ECEC centers.

FIGURE 1 The research design of the VUOKKO follow-up study.



1.1. Funding

The Toddler Phase (2015–1016, T1 and T2) of the VUOKKO study was funded by the Tiina and Antti Herlin Foundation (Jenni Salminen's postdoctoral funding), and the Faculty of Education and Psychology at the University of Jyväskylä, through research funding intended for hiring assisting research personnel, material costs or other support. The preschool phase (2018–2019, T3 and T4) was funded with financial support from the Academy of Finland (Jenni Salminen's postdoctoral funding No. 307080), the Emil Aaltonen Foundation (Grant for Jenni Salminen for covering data gathering expenses, 2019), and the Faculty of Education and Psychology at the University of Jyväskylä through research funding intended for hiring assisting research personnel, material costs or other support.

Data gathering for Grade 1 (2020–2021, T5) was funded by the Neo-PRISM-C project (www.neoprismc.org), an action of the Marie Sklodowska-Curie Innovative Training Network (ITN, Grant Agreement no. 813546) and the Faculty of Education and Psychology at the University of Jyväskylä through research funding intended for hiring assisting research personnel, material costs or other support.

Data gathering for Grades 2 and 3 (2021–2023, T6 and T7) was funded by the EarlyMath research project. EarlyMath has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement No 101002966) and by the CRITICAL research project funded by the Strategic Research Council (SRC) established within the Research Council of Finland (Grants #335625 and #335727).

Data gathering will continue in 2024–2025 in Grades 4 and 5 in the CRITICAL research project, which is funded by the Strategic Research Council (SRC) established within the Research Council of Finland (Grants #358490 and #358250).

1.2. Publications

A list of publications and theses in which the follow-up data were used and/or described can be found on the website of the VUOKKO study: www.jyu.fi/vuokko

1.3. Creators

PRINCIPAL INVESTIGATORS AND RESEARCH GROUPS

Toddler Phase (2015–2016, T1–T2)

- Principal investigators (PIs):
  - Marja-Kristiina Lerkkanen and Jenni Salminen

- The core research group (in alphabetical order):
  - Marja-Leena Laakso, Marja-Kristiina Lerkkanen, Eija Pakarinen, Anna-Maija Poikkeus, and Jenni Salminen

- International research groups:
    - The data for the VUOKKO study have been collected as part of the Quality Matters project, with each country holding ownership of the national dataset.

    - Quality Matters researchers:
        - Portugal (coordination and funding): Joana Cadima, Carolina Guedes, Teresa Aguiar, Cecilia Aguiar, and Clara Barata
        - Finland: Marja-Kristiina Lerkkanen and Jenni Salminen
        - Poland: Olga Wyslowska
        - The Netherlands: Pauline Slot

    - Collaboration with the University of Potsdam, Germany, to aid German researchers in the collection of the pilot dataset using measures similar to those in the VUOKKO study. Each country holds ownership of the national dataset.

    - Researchers involved in collaboration:
        - Germany: Gerlind Grosse and Katharina Nösser (Vielhaber)
        - Finland: Marja-Kristiina Lerkkanen and Jenni Salminen

Preschool Phase (2018–2019, T3–T4)

- Principal investigators (PIs):
    - Marja-Kristiina Lerkkanen and Jenni Salminen

- The core research group (in alphabetical order):
    - Tuire Koponen, Marja-Kristiina Lerkkanen, Eija Pakarinen, and Jenni Salminen.

School Phases, Grades 1–3 (2021–2023, T5–T7)

- Principal investigators (PIs):
    - Jenni Salminen, Marja-Kristiina Lerkkanen, Tuire Koponen, and Minna Torppa

- The core research group (in alphabetical order):
    - Kenneth Eklund, Daria Khanolainen, Tuire Koponen, Marja-Kristiina Lerkkanen, Tiia Nurminen, Eija Pakarinen, Jenni Salminen, and Minna Torppa

Data Collectors, Data Organizers and SPSS Inserters

Toddler and Preschool Phases (2015–2019, T1–T4)

- Coordinator:
    - Jenni Salminen

- Data collectors and data inserters:
    - Maija Koikkalainen, Christina Mantsinen, Jonna Ahvenkoski, Soile Runtti, Milja Parviainen, Jenni Salminen, Niina Matilainen, Viivi Öhman, Kati Vasalampi, Sini Ekroos, Essi Juhala, Emmi Mörsky, Satu Anttila, Kaisa Koski,

Tiia-Liina Raittila**,** Regina Egeler, Katariina Mäkelä, Elli-Maria Kamula, Viola Soininen.

School Phases, Grades 1–3 (2021–2023, T5–T7)

- Coordinators:
  - Essi Juhala (Grade 1, T5), Sanna Niemi (Grade 2, T6), Matilda Hamara (Grade 3, T7), Tiia Nurminen (Project coordinator: Parental assessment T5–T6, Individual child assessment Grade 3, T7).

- Data collectors and data inserters:
  - **T5**: Essi Juhala, Sonja Virtanen, Henni Elomaa, Tiia Nurminen, Marjukka Mäki-Kojola, Matilda Polojärvi, Elina Haarala, Sara Frimodig, Katriina Hakamäki, Emilia Liikala, Aliisa Manninen, Heidi Maaranen, Sami Pokka
  - **T6**: Sanna Niemi, Matilda Hamara, Marjukka Mäki-Kojola, Mesi Katisko, Tanja Tanskanen, Tiina Mäki-Kojola, Soile Runtti, Tarja Honkanen, Leena Sorvari, Juulia Kytömäki, Henni Elomaa, Kyösti Sorsa, Suvi Heikkinen, Sara Puustelli, Merika Kallio, Anna Töllinen, Tiia Nurminen, Jenni Salminen
  - **T7:** Matilda Hamara, Tiia Nurminen, Anniina Heikkinen, Kaisa Tapanainen, Nea Nurmisto, Susanna Perttilä, Vappu Koho, Venla Bom, Lotta Sieppi, Nea Hannula, Tiia Inkilä, Juulia Kotiranta, Tatu Hoffren, Sanni Heikkinen, Juho Mäki, Johanna Passilahti, Sofia Fabritius, Tiina Lehtonen, Sini Lehtinen, Johanna Mehtonen, Jenni Jokikuona, Maiju Mykrä, Emma Roikonen, Markus Kaari, Irina Pitkänen, Nea Hannula, Santeri Toikkanen
  - **T5–T6, Parental assessment**: Tiia Nurminen, Essi Juhala, Jarno Rautiainen, Marjukka Mäki-Kojola, Alisa Salminen, Matilda Hamara, Mesi Katisko, Tanja Tanskanen, Tiina Mäki-Kojola, Inka Piirainen, Touko Torppa

1.4. Access to Data

The VUOKKO research group welcomes research collaboration and wishes to contribute to the research according to the Findable, Accessible, Interoperable, Reusable (FAIR) principles (see e.g., Wilkinson et al., 2016). At the time of writing this metadata file, the data are located on the secured servers of the University of Jyväskylä. Access to the data is coordinated by a steering group and the researchers interested in the data will receive further guidelines for data access from the data contact person. The steering group coordinates publishing and takes care of contracts for data use (IPR, data protection), information flow within the research group, and access to the data.

2. Ethical Considerations

The VUOKKO study has followed the research ethics guidelines set out by ALLEA (2017), the Finnish National Board on Research Integrity (2012), the World Medical Association Declaration of Helsinki (WMA, 2013), and the University of Jyväskylä (JYU): (https://www.jyu.fi/en/research/responsible-science/human-sciences-ethics-committee).
VUOKKO has complied with the EU General Data Protection Regulation (GDPR) and other legislation in question and relevant during the data-gathering timepoints. The research has met

the national legal and ethics requirements. This means that VUOKKO has complied with GDPR 2016/679 Directive 95/46/EC (the protection of individuals with regard to the processing of personal data and on the free movement of such data). VUOKKO has also complied with national legislation, including legislation for the protection of privacy (Data Protection Act (1050/2018) (Tietosuojalaki 1054/2018). The ethics risks related to the data processing activities of the project have been evaluated according to JYU guidelines throughout the follow-up (https://tietosuoja.fi/en/impact-assessments).

In all phases of the research project, the research group members, students and research assistants have committed themselves to acting for the welfare of the participants.

Ethical Evaluation

The data collection for the Toddler Phase of the VUOKKO study was part of the joint research project Quality Matters between four European Countries (Finland, Poland, Portugal and the Netherlands). As part of the preparation stage, the recruitment and data collection protocols (similar across all countries) were pre-evaluated by the Portuguese National Commission of Data Protection, and a positive statement was received. For the Finnish national data collection, the ethical pre-evaluation by the Ethics Committee of the University of Jyväskylä was not deemed necessary when the ECEC data collection timepoints began in 2015 and in 2018, based on the guidelines of the Finnish National Board on Research Integrity (TENK), which the University of Jyväskylä has undertaken to comply with (https://www.jyu.fi/en/research/responsible-science/human-sciences-ethics-committee/tarvitseeko-tutkimuksesi-eettista-ennakkoarviointia). This was ensured afterwards by asking the Ethics Committee for an official statement on the need for an ethical review. Based on the research plan, the Ethics Committee chair and secretary assessed the issue, and concluded that an ethical review and statement by the Committee was not needed for the ECEC Phase of the VUOKKO study (#352/13.00.04.00/2023).

Ethical pre-evaluation for the School Phase was sought and received from the Ethical Committee of the University of Jyväskylä in spring 2020, before the follow-up children went to Grade 1 (#613/13.00.04.00/2020). This pre-evaluation covered the three data collection timepoints in primary education (Grades 1–3).

Participant Consent

The VUOKKO study has involved assessments of human participants, and thus informed consent has been obtained for ethical reasons. All information sheets and consent forms were written in clear language.

Children's consent: The legal guardians were asked for consent for the children's participation. After school entry, the children who could write their name were also asked to sign a consent form, which were written as clearly and concisely as possible. Before the actual assessments, the researchers made sure that the children had an opportunity to ask questions about the research and were not forced to participate. Researchers allowed the children to discontinue the assessment if the children were not willing to continue. Researchers were instructed to practice special sensitivity with the children because they may not always be able to communicate with words about their willingness to participate.

Parents and teachers gave consent concerning their own participation in the study. We ensured that participants fully understood the information given to them about the study and did not feel pressured or coerced into giving consent.

Information Sheets

To ensure informed consent, we provided information sheets together with the consent forms. The legal guardians of the participating children were asked to discuss the participation details with the children to ensure their voluntary participation. In the information letter, the contact details of the research group were given, and the legal guardians were encouraged to contact the research group in the case of any further questions. For children who were able to read (in our sample in Grades 1–3, at the age of 7–10 years), we also provided a separate, easier information sheet that took into account the age of the child.

The following issues were explained clearly and in everyday language in the information sheets:

1. Contact information of the researchers and the principal investigator.
2. Project background information and research institute.
3. Purpose, objective, and scientific and practical implications of the research.
4. Methods and procedures to be used with the participants.
5. The nature of the participation.
6. Any benefits, risks, or discomfort that might ensue.
7. How and for what purposes the acquired information will be used.
8. How and where data will be collected. How much time participation is expected to take.
9. How data will be protected during the project and either destroyed or reused subsequently, including the method of data storage that ensures it cannot be accessed by anyone outside the research group and the method of securing the identity of the participants.
10. The participants' rights: they can refuse to participate in the research; they can, in any phase, ask for additional information on the research; and they can cancel their participation in the research in any phase without consequences.
11. Participants' insurance coverage.

Privacy Notice

When the participants were asked for their consent to participate, they were given a privacy notice which notified them about how the privacy issues are handled in VUOKKO. We followed the University of Jyväskylä guidelines for data privacy (https://www.jyu.fi/en/university/data-privacy) and utilised the available JYU template of the privacy notice (https://www.jyu.fi/en/university/data-privacy/data-privacy-templates) for drafting the project's own privacy notice for each data collection timepoint, which was shared for the participants through the website for the VUOKKO study.

Ethical Principles Leading the Training of Research Assistants

Training of the research assistants is the most central way of ensuring the ethical conduct of research in practice. Prior to completing any child or parental assessments or classroom observations, all research assistants attended training during which the key ethical principles of the project were discussed with them. These principles covered the following:

- ***Confidentiality and responsible conduct of research***, including the secure handling of any personal data, paperwork and assessment sheets.
- Encountering the children, their parents and teachers during the data collection
- ***Voluntariness and non-pressurized participation of the study informants***. For example, observing the young children's willingness to participate, communication within the classrooms or with the parents.
- ***Informed consent.*** Only participants with consent took part in the assessments and/or observations. Nevertheless, to maintain fair and equal treatment for all children in schools, in most classrooms all children participated in tasks regardless of having a consent for that. For those children without the consent, the sheets were taken to university and disposed in a secure way. None of these sheets were scored for the analysis.
- Dealing with ***incidental findings.*** Incidental findings are observations of potential clinical significance unexpectedly discovered among the participants, and unrelated to the purpose of the study. The training of the researchers and research assistants included instructions on how to handle these types of situations to ensure that each researcher and research assistant was prepared for such incidences while visiting the ECEC settings and schools and meeting the parents attending the study.
- ***Personal well-being of the research assistants***, e.g., collegial support from research team members and other research assistants.

Since August 2021, in accordance with the requirements of the EarlyMath funder, the European Research Council, all researchers and research assistants conducting data collection with minors at schools were asked to present an up-to-date ***criminal records check*** provided by the legal register center. The check needed to be less than 6 months old, and it had to be requested either for pedagogical/educational studies (BA and MA students) or for scientific work with minors (hired staff). The record checks were shown to the responsible researcher during the given data collection period.

3. Data Management Principles

Data have been collected over several years during the VUOKKO study, from fall 2015 to spring 2023. Table 1 introduces the types of data gathered. For the timeline of the data collection, see Figure 1.

Data Protection

Data protection has been carefully considered from the early stages of the research project onwards, by applying the strictest privacy settings throughout the project's lifecycle, and all actions in the project have complied with respective national legal frameworks.

In data management and data security we have followed the current and updating guidelines of the University of Jyväskylä. The current guidelines can be found here: https://www.jyu.fi/en/university/data-privacy/tietosuojaohjeet/researchers.

Since 2021, when the university-level courses became available, all researchers have been required to complete the courses provided by the University of Jyväskylä on data privacy issues and on data protection.

TABLE 1 Overview of the Data Types, Collection Methods and Data Formats

Content and type of data collected	**Collection method**	**Data format**
Assessments of children's literacy and numeracy skills and related cognitive skills. Performance is scored in numerical form.	Group and individual assessments / psychological tests.	Excel, SPSS
Children's questionnaires on their interest in reading and math and reading and math anxiety.	Paper questionnaires	SPSS
Assessments of parents' literacy and numeracy skills and related cognitive skills. Performance is scored to numerical form.	Individual assessments	SPSS
Parental questionnaires on their children (e.g., numerical and literacy development, play interests, temperament) and themselves (e.g., education and occupation, self-concept of mathematical and reading skills, attitudes towards math and reading) and the activities in the home environment (e.g., frequency of shared reading or engaging in informal and formal numeracy activities).	Paper questionnaires / Web-based questionnaires via Webropol	Excel, SPSS
Observational data in the ECEC centers and primary school classrooms. Video recordings are coded according to a specific coding scheme yielding numerical data.	Video and/or audio recordings, field-notes	Excel, SPSS, video recording (mp4/MTS), audio recording (mp3/WMA). Text files for transcripts.
Teachers' questionnaires on themselves and their pedagogical work (e.g., education, views on implementing pedagogy, well-being, working conditions, self-efficacy beliefs).	Paper questionnaires	SPSS
Teachers' reports concerning individual children (e.g., self-regulation, social skills, individualization of instruction).	Paper questionnaires	SPSS

Agreements

Since the beginning of the study, all students, before receiving data from the VUOKKO study for their BA/MA/PhD theses, or starting to work at the data collecting tasks, have signed an agreement covering the details of ***secure handling of research data*** ('datankäyttösopimus'), which has been prepared by the faculty of Education and Psychology for use in research projects.

Since August 2021, in accordance with the requirements of the EarlyMath funder, European Research Council, all researchers and research assistants (including BA and MA students taking part on data collection, organizing or inserting the data**)** have signed three contracts/ agreements: (1) ***Agreement for use of research data*** ('datankäyttösopimus'), (2) ***Agreement of processing personal data*** during the research study ('tietojenkäsittelysopimus') and (3) ***transfer of rights agreement*** ('oikeuksiensiirtosopimus'). All the signed documents have been scanned and sent to the JYU Registry Office and Archives.

Data Location

All VUOKKO study follow-up data is stored on university servers and require passwords according to the University of Jyväskylä's up-to-date guidelines. The data that have been collected on paper (e.g., questionnaires and assessment sheets) have been scanned and stored to the university servers and the papers have been securely disposed of. The video and audio recordings are stored securely at the university servers. No other persons than researchers of this research group have access to the data, and the access requires login with a university username and password. However, access to the dataset can be granted through the steering groups of the VUOKKO study (2015–2019: Lerkkanen & Salminen, 2020–2023: Salminen, Lerkkanen, Koponen & Torppa).

Data Pseudonymization

The data is pseudonymized by assigning each participant a numerical code, which is then used in the data and in all analyses. Data have also been collected by using electronic encoding tools (digital recorders) and video recordings, which are given special attention. The voice recordings have been gathered to support scoring of the individual skill assessments. These recordings have been deleted as soon as possible once the data have been coded in numerical format. The data from video-recorded interactions have been coded from the recordings into numerical form (CLASS tool) according to the up-to-date guidelines of the University of Jyväskylä. Direct identifiers, however, cannot be erased from the original audio or video data. The original audio and video files are stored on secure university servers and only the members of the research group can gain access to them.

4. Data Collection

4.1. RP – Recruitment and Participants

4.1.1. RP – Toddler Phase (2015–2016, T1–T2)

Participants

The Toddler Phase of the VUOKKO study (2015–2016) comprised a sample of 245 children in 43 ECEC classrooms (spread across 36 ECEC centers), their parents (N = 208) and ECEC teachers (N = 117) (See Figure 1). The sample was drawn from the ECEC settings in the city of Jyväskylä (including public, non-profit ECEC centers from rural and urban areas) as a joint

development project with the city's ECEC services. The data were ultimately gathered as part of the Quality Matters project,[4] in which researchers in four European countries agreed to collect a comparable sample of research data on teacher–child interaction, teacher and classroom characteristics, and child outcomes across 30 ECEC centers in each country.

The study began in September 2015 with children born in 2013 (aged 2 to 3 years) and who attended ECEC in Jyväskylä.

Recruitment and Consent Protocols

Data gathering began in October 2015 in the city of Jyväskylä. The head of the ECEC services (responsible for research and development) was consulted in August 2015 and a joint agreement on collaboration between the city and university was established and informed consent was granted by the city of Jyväskylä. As part of the collaboration, at least one classroom from each ECEC center within the city serving children between the ages of 2 and 3 (born in 2013) was expected to participate in the study. Participation was nevertheless based on voluntarism. There were altogether 53 public, non-profit ECEC centers in the city of Jyväskylä in the fall 2015, spread across rural and urban areas. Based on the age range of the target children in the current study, 43 ECEC centers and 51 classrooms within them were identified as suitable to participate. Out of these, seven ECEC centers and 8 classrooms within them declined to participate. Teachers representing the participating ECEC centers (N = 36) and classrooms (N = 43) were invited to a joint kick-off meeting (September 2015; 2 hr) during which the timeline of the research was introduced, and the data collection methods were discussed together with the teachers. At the same meeting, the consent forms for ECEC teachers and parents were given to the teachers, who took the forms with them and delivered them to the ECEC classroom teachers and to the parents of the children in their own ECEC classrooms. In case the teacher was not able to participate in the kick-off meeting, the consent forms were sent through mail directly to the ECEC centers.

The consent forms were delivered to the parents via ECEC centers and parents returned the forms to the teachers. Teachers returned the consent forms in one big envelope as soon as all the consent forms from the parents were returned. Teachers also filled in the consent form regarding their own participation (each teacher in the ECEC classroom, range 1–5). Within this study, (due to highly heterogeneous age groups in Finnish ECEC settings), all children born in 2013 were included in the study (taking part in child assessments, teacher evaluations and observations) and all the other children in the classrooms (older/younger) were included to be present during the video recordings. Consent forms were provided in Finnish for the follow-up children and their families. Due to the strong role of the Finnish language used in the individual skill assessment, participation in the study was mainly restricted to Finnish-speaking children. However, the consent forms concerning the video recording participation were translated additionally into English, Swedish, Russian, Swahili, Kinyarwanda, Persian, and Kurdish to correspond to the varying home languages of the children in the participating ECEC classrooms.

[4] The Quality Matters project included researchers from Portugal, Poland, Netherlands, and Finland. The research collaboration was funded by the Portuguese Foundation for Science and Technology (grant PTDC/MHC-CED/5913/2014, research project "Quality Matters").

Throughout the years of the study there were changes in the classrooms as new children arrived and others moved to other classrooms. Therefore, new consent forms were delivered to the ECEC centers when needed. New children were thus included in the sample throughout the year: both the VUOKKO follow-up children (born in 20T13) and their classmates, who only took part in the video recording. Follow-up of the children who changed classrooms was discontinued unless the receiving classroom was also participating in the VUOKKO study. Some of these children changed classrooms within the city, some moved away from Jyväskylä, and some stopped attending ECEC. The reason for withdrawal from ECEC for children aged 2 to 3 is typically that there is a new baby born in the family and parents take care of both children at home.

There were 485 children born in 2013 who were attending municipal ECEC in Jyväskylä in September 2015. The number of children eligible to be invited to the VUOKKO study in the eligible ECEC centers (N = 43) in the fall 2015 was 322. Due to drop out of eight ECEC classrooms and not all parents giving their consent, the final number of children participating in the VUOKKO sample in the toddler year was 264. Eventually, the skill assessment data were available for 245 children (fall of the toddler year N = 230 and in the spring N = 205). Of these children N = 190 participated both in fall (T1) and in the spring (T2). 54% of the children born in 2013 who were taking part in municipal center-based ECEC in Jyväskylä (excluding children in private ECEC, clubs or family-based care), participated in the VUOKKO study.

Teachers in 41 out of 43 participating classrooms agreed to be part of the classroom observation and at least one teacher in each of the 43 classrooms agreed to respond to the teachers' questionnaire and complete the report on individual children.

Individual teachers did not receive personal gifts for participating in the study, but a children's book (*Oona ja Eetu – Päivä hoidossa*) was brought to each child group after completing the observation.

4.1.2. RP – Preschool Phase (2018–2019, T3–T4)

Participants

The Preschool Phase of the VUOKKO study (2018–2019) continued in August 2018 with the same children (born in 2013, now age 5 to 6) who had participated in the study as toddlers. The Preschool Phase consisted of a sample of 189 children in 61 ECEC classrooms (spread across 32 ECEC centers), their parents (N = 134) and ECEC teachers (N = 98) (See Figure 1). The sample was drawn from the ECEC settings in the city of Jyväskylä (including public, non-profit ECEC centers from rural and urban areas).

Recruitment and Consent Protocols

Informed consent was first sought from the City of Jyväskylä. The recruitment process then continued by asking for a list of names of all children born in 2013, currently attending ECEC in Jyväskylä, from the administration of the city's ECEC. The name list was then used to identify the current ECEC classrooms of the VUOKKO follow-up children and their current ECEC teachers. These teachers were invited to a joint kick-off meeting (late August 2018, 2 hr) during which the timeline of the follow-up design of the research was introduced, and the data collection methods were discussed together with the teachers. At the same meeting

teachers could take the informed consent forms. A large majority of the teachers were unable to attend the kick-off meeting, and these teachers were later contacted via phone calls to inquire about their interest to take part in the follow-up study. For these teachers and the children in their classrooms, the informed consent forms were delivered via mail to ECEC centers.

The consent forms were delivered to the parents via ECEC centers and parents returned the forms to the teachers. Teachers returned the consent forms in one big envelope as soon as all the consent forms from the parents were returned. Teachers also filled in the consent form regarding their own participation (each teacher in the classroom). Within the VUOKKO study (due to highly heterogeneous age groups in Finnish ECEC settings), all children born in 2013 were included in the study (taking part in child assessments, teacher evaluations and observations) and all the other children in the classroom (older/younger) were included to be present during the video recordings. Due to the follow-up nature of the VUOKKO sample, consent forms were provided in Finnish only. The consent forms concerning the video recording participation were translated additionally into English.

There were 1,020 children born in 2013 who were attending municipal ECEC in Jyväskylä in September 2018. The name lists were skimmed in the search for the 264 VUOKKO follow-up children who had taken part in the Toddler Phase of the study. We were able to identify and locate 220 children who had participated in the Toddler Phase in 2015 to 2016. At the same time, 39 children, who had participated in the Toddler Phase of the VUOKKO study, were no longer accessible through municipal ECEC in Jyväskylä and their participation in the study was discontinued. During the recruitment process, 20 (out of 52) ECEC centers in Jyväskylä decided not to take part in the Preschool Phase, and due to this, 29 children were no longer included in the study.

In 12 classrooms, in which there was only one VUOKKO follow-up child, who was born in 2013 and had attended the study in the Toddler Phase, another child born in 2013 was recruited from that classroom and included in the follow-up sample in the preschool phase. This added 12 new follow-up children to the sample. Consequently, consent forms were sent to the parents of 208 children.

The final number of children in the preschool phase sample was 191. Eventually, the skill assessment data were available for 189 children (fall of the preschool year N = 188, and in the spring N = 175). Of these children N = 174 participated in both in the fall (T3) and in the spring (T4). Of the families who were invited, 92% participated.

Teachers in all 61 participating classrooms agreed to be part of the classroom observation (5 classrooms were observed without video recording) and at least one teacher in each of the 61 classrooms agreed to respond to the teachers' questionnaire and complete the report on individual children.

Individual teachers did not receive personal gifts for participating in the study, but a children's book (*Molli ja Kumma*) was brought to each child group after completing the observation. In addition, after completing the data collection a prize draw was held among all participating teachers and five gift vouchers (each worth €50) were awarded at random.

4.1.3. RP – School Phase: Grade 1 (2020–2021, T5)

Participants

The VUOKKO study in Grade 1 (2020–2021) comprised 666 children in 50 primary school Grade 1 classrooms, their teachers (N = 40) and parents (N = 446) (See Figure 1). Of the 666 participating children, 151 were VUOKKO follow-up children. The sample was drawn from one Finnish city (including public primary schools from rural and urban areas). The study continued in Grade 1 with the same children (born in 2013, now age 7 to 8) who had participated in the study in ECEC centers, and their classmates.

NOTE! Data collection took place during the COVID-19 pandemic. Due to health risks, the originally planned data collection from Grade 1 in fall 2020 was fully cancelled and the data collection only took place in spring 2021. The recruitment process, described below, took into account the COVID-19 situation in every school and the necessary amendments to the data collection were made at the school level. Two schools declined to participate in the study due to having gone through several quarantines and experiencing overall challenges in terms of the COVID-19 situation.

Recruitment and Consent Protocols

Informed consent was first sought from the City of Jyväskylä. The research plan and permit application were delivered to the representative in charge of study permits in the City of Jyväskylä. The recruitment process then continued by asking for a list of all children born in 2013, currently attending primary education in Grade 1 classrooms in Jyväskylä, from the secretary responsible for personal information systems. Based on the name list, the current classrooms and teachers of the VUOKKO follow-up children were identified. Altogether, 236 children who had participated in some or several earlier phases of the VUOKKO follow-up were found from 26 different schools and 63 classrooms. There were altogether 1,209 children in the classrooms that also had at least one VUOKKO-follow-up child.

Starting in mid-December 2020, the school principals were contacted by phone and email to explain the nature of the VUOKKO study. After this, either the rector or the research coordinator contacted the teachers and inquired about their interest to take part in the study. Teachers were informed about the aim of the study as well as the timeline and the data collection methods. Each teacher received research information via email, with an attachment explaining the study designs and details of data collection principles. In the emails, teachers were provided with a possibility for a phone call in which the details of the study could be further discussed. Some teachers indicated that the phone call would not be needed and declined their participation at this stage, but the majority of the teachers wanted to discuss on the phone. Teachers had a generally positive stance toward participation, mainly due to the fact that the planned data collection was part of an ongoing longitudinal study. In case the teachers were not willing to participate in the study in its full scale, further negotiation was carried out: Teachers were provided with a possibility to only allow children's skill assessments in the classrooms and to choose which other parts of the data collection they would be willing to take part in. For instance, some teachers declined classroom observation, but were willing to fill in the questionnaires.

Altogether, 50 teachers (out of 63) agreed to be part of the study by allowing children's assessment in their classrooms. Of these, 42 teachers were willing to respond to questionnaires

and 40 of them replied. There were 43 teachers willing to fill in the individual assessment of their students and 41 returned the evaluations. And, finally, 37 teachers were willing to participate in the classroom observation, and eventually classroom observations were conducted in each of these classrooms.

In these 50 classrooms, there were altogether 947 children, of which 175 had participated in the VUOKKO study in some or several earlier phases. These children and their families were contacted and invited to participate in the study in January 2021.

Informed consent forms were delivered directly to the schools in January 2021 by the research coordinator and they were distributed to the families by the class teachers (children brought the consent forms home to their parents). Consent was requested from only one parent from each family. The parents returned the consent forms, via their children, to the teachers, who then sent the completed forms to the research coordinator through the mail. The consent forms were provided mainly in Finnish, but teachers could request an English translation from the researchers if needed for some families. These translated versions were sent to two classrooms. Additionally, for some families in three classrooms a simplified Finnish version of the consent form was provided, in which the participation and key ethical aspects of the study were explained with brief sentences and simplified language.

Despite the teachers' efforts, not all families returned the consent forms actively to schools: Of the 947 families invited, 712 families gave consent for their child to participate in the study. Eventually, the skill assessment data were available for 666 children. Out of these children, 151 had participated in the VUOKKO study earlier in the Toddler and/or in Preschool Phases. The main reason for the lower number of children assessed than children participating in the study was the overall COVID-19 situation in the schools during the data collection. Children were kept home if they had any flu symptoms, which meant that several children might have been absent during the assessment day and the researchers were not always able to complete the tasks with all children in the classrooms. Only in cases when the VUOKKO follow-up child was absent would the researcher return to school and complete the assessment with the child. All other back-up visits were refrained from so as to avoid multiple visits to classrooms during the COVID-19 pandemic.

Individual teachers received €20 gift vouchers for participating in the study. Children received small rewards (a sticker) after completing the assessment in the classroom. In addition, after completing the data collection a prize draw was held among all participating teachers and one gift voucher (worth €50) was awarded at random.

4.1.4. RP – School Phase: Grade 2 (2021–2022, T6)

Participants

The VUOKKO study in Grade 2 (2021–2022) comprised 659 children in 48 primary school Grade 2 classrooms, their teachers (*N* = 42) and parents (*N* = 370) (See Figure 1). Of the 659 participating children, 151 were VUOKKO follow-up children. The sample was drawn from one Finnish city (including public primary schools from rural and urban areas). The study continued with the same children (born in 2013, now age 8 to 9) who had participated in the study in ECEC centers as well as in Grade 1 (2020–2021).

NOTE! Similar to the Grade 1 data collection, the Grade 2 data collection took place during the COVID-19 pandemic. Due to the national pandemic situation blocking the outsiders' visit to school during January and February 2022, data collection was postponed toward the beginning of March 2022. The recruitment process, described below, took into account the COVID-19 situation in every school and the necessary amendments to the data collection were made at the school level. Three schools (seven classrooms) declined to participate in the study due to having gone through several quarantines and experiencing overall challenges due to the ongoing pandemic, but also due to other simultaneous burdening factors.

Recruitment and Consent Protocols

Informed consent was first sought from the City of Jyväskylä. Informed consent was sought for two subsequent years at the same time, including data collection when the VUOKKO follow-up children were in Grade 2 (2021–2022) and Grade 3 (2022–2023). The recruitment process then continued by asking for a list of all children born in 2013, currently attending primary education in Grade 2 classrooms in Jyväskylä, from the city administration (responsible for basic education). Based on the name list, the current classrooms and teachers of the VUOKKO follow-up children were identified, and the participation list from Grade 1 was compared to the current name list.

In December 2021 and partially in January 2022, the school principals were contacted by phone or email to explain the nature of the VUOKKO study and inquire about the possibility to continue follow-up with Grade 2 classrooms and teachers. After this, either the rector or the research coordinator contacted the teachers and inquired about their interest to take part in the study. Teachers were informed about the aim of the study as well as the timeline and the data collection methods. Each teacher received research information via email, with an attachment explaining the study designs and details of data collection principles. In the emails, teachers were provided with a possibility for a phone call in which the details of the study could be further discussed. Some teachers indicated that the phone call would not be needed and declined their participation at this stage, but the majority of the teachers wanted to discuss on the phone. Teachers had a generally positive stance toward participation, mainly due to the fact that the planned data collection was part of an ongoing longitudinal study. In case the teachers were not willing to participate in the study in its full scale, further negotiation was carried out: Teachers were provided with a possibility to only allow children's skill assessments in the classrooms and to choose which other parts of the data collection they would be willing to take part in. For instance, some teachers declined classroom observation, but were willing to fill in the questionnaires.

Altogether, 48 teachers (out of 65) agreed to be part of the study by allowing children's assessment in their classrooms. Of these, 41 teachers were willing to respond to questionnaires and 40 of them replied. There were 38 teachers willing to fill in the individual assessment of their students and 36 returned the evaluations. And, finally, 36 teachers were willing to participate in the classroom observation, and eventually classroom observations were conducted in each of these classrooms.

In these 48 classrooms, there were altogether 952 children, of which 198 had participated in the VUOKKO study in the Toddler and/or in Preschool Phases. All these children and their families were contacted and invited to participate in the study in January 2022.

Informed consent forms were delivered directly to the schools in January 2022 by the research coordinator and they were distributed to the families by the class teachers (children brought the consent forms home to their parents). Consent was requested from only one parent from each family. The parents returned the consent forms, via their children, to the teachers, who then sent the completed forms to the research coordinator through the mail. The informed consent covered both Grade 2 and 3 for children and their families. For teachers, the consent was collected for each grade level separately as it was expected that several teachers might change on the verge of Grade 3. The consent forms were provided mainly in Finnish, but teachers could request an English translation from the researchers if needed for some families. Additionally, for some families a simplified Finnish version of the consent form was provided, in which the participation and key ethical aspects of the study were explained with brief sentences and simplified language. This type of consent was delivered to five families within one classroom.

Despite the teachers' efforts, not all families returned the consent forms actively to schools: Of the 952 families invited, 688 families gave consent for their child to participate in the study. Eventually, the skill assessment data were available for 659 children. Out of these children, 151 had participated in the VUOKKO study earlier in the Toddler and/or in Preschool Phases. One reason for the lower number of children assessed than children participating in the study was the overall COVID-19 situation in the schools during the data collection. Children were kept home if they had any flu symptoms, which meant that several children might have been absent during the assessment day and the researchers were not always able to complete the tasks with all children in the classrooms. Only in cases when the VUOKKO follow-up child was absent would the researcher return to school and complete the assessment with the child. All other back-up visits were refrained from so as to avoid multiple visits to classrooms during the COVID-19 pandemic.

Individual teachers received €20 gift voucher for participating in the study. Children received small rewards (a sticker) after completing the assessment in the classroom.

4.1.5. RP – School Phase: Grade 3 (2022–2023, T7)

Participants of the Group Assessments

The VUOKKO study in Grade 3 (2022–2023) comprised 702 children in 55 primary school Grade 3 classrooms, their teachers (N = 45) and parents (N = 493) (See Figure 1). Of the 702 participating children, 161 were VUOKKO follow-up children. The sample was drawn from one Finnish city (including public primary schools from rural and urban areas). The study continued with the same children (born in 2013, now age 9 to 10) who had participated in the study in ECEC centers and in Grade 2 (2021–2022).

Recruitment and Consent Protocol of the Group Assessment

Informed consent was sought when the children were in Grade 2. The consent applied for the data collection that took place in both Grades 2 and 3. Hence, most of the children and their parents had already given consent for the data collection for Grade 3. Nevertheless, six new classrooms that had not participated in the study during Grade 1 or Grade 2 were willing to participate in Grade 3. Furthermore, four classrooms that had been part of the study in the

Grade 1 but not in the Grade 2 were willing to participate in Grade 3. And, in addition, three classrooms that had participated in the study in Grade 2 were no longer willing to participate in the study in Grade 3. The consent protocol and recruitment were conducted with those classes according to the same principles as in Grade 2. See more detailed information on the recruitment process and consent protocol in section 4.1.4.

Altogether, 55 teachers (out of 70) agreed to be part of the study by allowing children's assessment in their classrooms. Of these, 54 teachers were willing to respond to questionnaires and 45 of them replied. There were 49 teachers willing to fill in the individual assessment of their students and 41 returned the evaluations. And, finally, 43 teachers were willing to participate in classroom observation, and eventually classroom observations were conducted in 39 classrooms.

In these 55 classrooms, there were altogether 1,055 children, of which 186 had participated in the VUOKKO study in the Toddler and/or in Preschool Phases. All these children and their families were contacted and invited to participate in the study either in January 2022 or in January 2023.

Despite the teachers' efforts, not all families returned the consent forms actively to schools: Of the 1,055 families invited either in Grade 2 or Grade 3, 752 families gave consent for their child to participate in the study. Eventually, the skill assessment data were available for 702 children. Out of these children, 161 had participated in the VUOKKO study earlier in the Toddler and/or in Preschool Phases.

In the cases when the VUOKKO follow-up child was absent, the research assistants returned to school and completed the assessment with the child. All other back-up visits were retained. With this procedure, the assessment was completed with all VUOKKO follow-up children with consent for participation but led to not receiving data from all children with consent. Furthermore, in cases when children's own consent sheet was empty (i.e., there was no information on the child's own consent), children's assessments were not used in the data. New consent forms were sent to these children's homes, but due to not receiving all of them back, these children's data were excluded from the final dataset.

Individual teachers received either a puzzle with a small book about Finland or a €20 gift voucher for participating in the study. Children received small rewards (a bookmark) after completing the assessment in the classroom.

Participants of the Individual Assessments

The individual assessment of the VUOKKO study in Grade 3 (2022–2023) comprised a subsample of 325 children from 55 classrooms, and 1 to 15 children participated per class.

Recruitment and Consent Protocol of the Individual Assessments

Of the 752 children who had consent to participate in the study, a subsample of 325 children was invited to participate in the individual assessment as well. All children were not assessed individually due to lack of sufficient resources. Three different entry criteria to the individual assessments were used in line with the project's research questions: First, the VUOKKO follow-up children, who had attended the study in ECEC centers, were invited because of the availability of the long-term longitudinal data. Second, the children who scored in the lowest

10th percentile in at least one of the three calculation fluency measures (Addition fluency & Subtraction fluency tasks' mean score, Arithmetic fluency task, and Number comparison task with multi-digit numbers) in Grade 2 were invited to oversample students with math difficulties. Third, the children, whose parents had participated in the parental assessment were also invited to maximize the amount of intergenerational data.

Eventually, the individual skill assessment data were available for 325 children. Out of these children, 162 had participated in the VUOKKO study earlier in the Toddler and/or in Preschool Phases, 112 had had math difficulties in Grade 2, and there was parental skill data from at least from one parent for 180 of them.

The children received a sticker for participating in the individual assessment.

4.1.6. RP – Parental Assessments (2021–2022, T5–T6)

Participants

Children's biological parents were invited to the Jyväskylä university facilities to participate in the assessments of their own mathematical and literacy skills. If possible, both parents were invited to participate in the study. Altogether 295 parents participated in the assessment, of which 166 were biological mothers and 129 were biological fathers. There were 79 families from which both parents participated. Of the participating 295 parents, 132 were parents of VUOKKO follow-up children and both parents participated in 41 families.

NOTE! Data collection took place during the COVID-19 pandemic. Due to this the researchers and participants were instructed to use face masks according to the rules of the JYU. If the parent could not wear a face mask, they were invited to participate in summer 2022, when there was no longer a recommendation to use a face mask.

Recruitment and Consent Protocols

Consent to participate in the parental assessment was requested when the children were in Grade 1 (in spring 2021, for a more specific description, see section 4.1.3.). Only one parent from each family was asked to fill in the consent form. Altogether 504 parents gave consent, of which 122 were the parents of VUOKKO follow-up children.

The recruitment process for the parental assessment started by calling to invite them to participate in the assessment. The exclusion criterion used for the participation was that if the guardian was not a biological parent of the child, they were not invited to participate in the assessment. If the parent was willing to participate in the assessment, the timetable was set on the phone, and they were given short instructions on how to come to the university facilities. After the phone call they were sent a text message that included the same brief information that was earlier agreed on the phone. A few days prior to the assessment day the parents received a new text message as a reminder about the upcoming assessment.

In the phone call, the parents were asked if the child's other biological parent is present in their life. If yes, they were asked whether they could give our phone number to that parent, or alternatively whether they could give the parent's phone number to us, so that we could try to reach them. If the other parent was also interested in participating in the assessment, the

recruitment process was conducted as described above. The only difference was that these parents were introduced to the study more specifically on the phone since it was possible that they had not read the information letter earlier. In the assessment they also had a chance to read the whole information letter and they signed the consent form before starting the assessment. Some of the children's other biological parents were only willing to fill in the parental questionnaire, but not to participate in the assessment at the university. In these cases, they were sent a parental questionnaire, information letter, and consent form and they mailed the completed questionnaire and consent form back to the university.

Out of the 504 parents that originally had given their consent, 271 were invited to the assessment. Of those parents, eventually 205 participated in the assessment, and additionally 90 children's other biological parents participated. Consequently, altogether 295 parents were assessed. Not all parents that had given consent were assessed due to resource limitations.

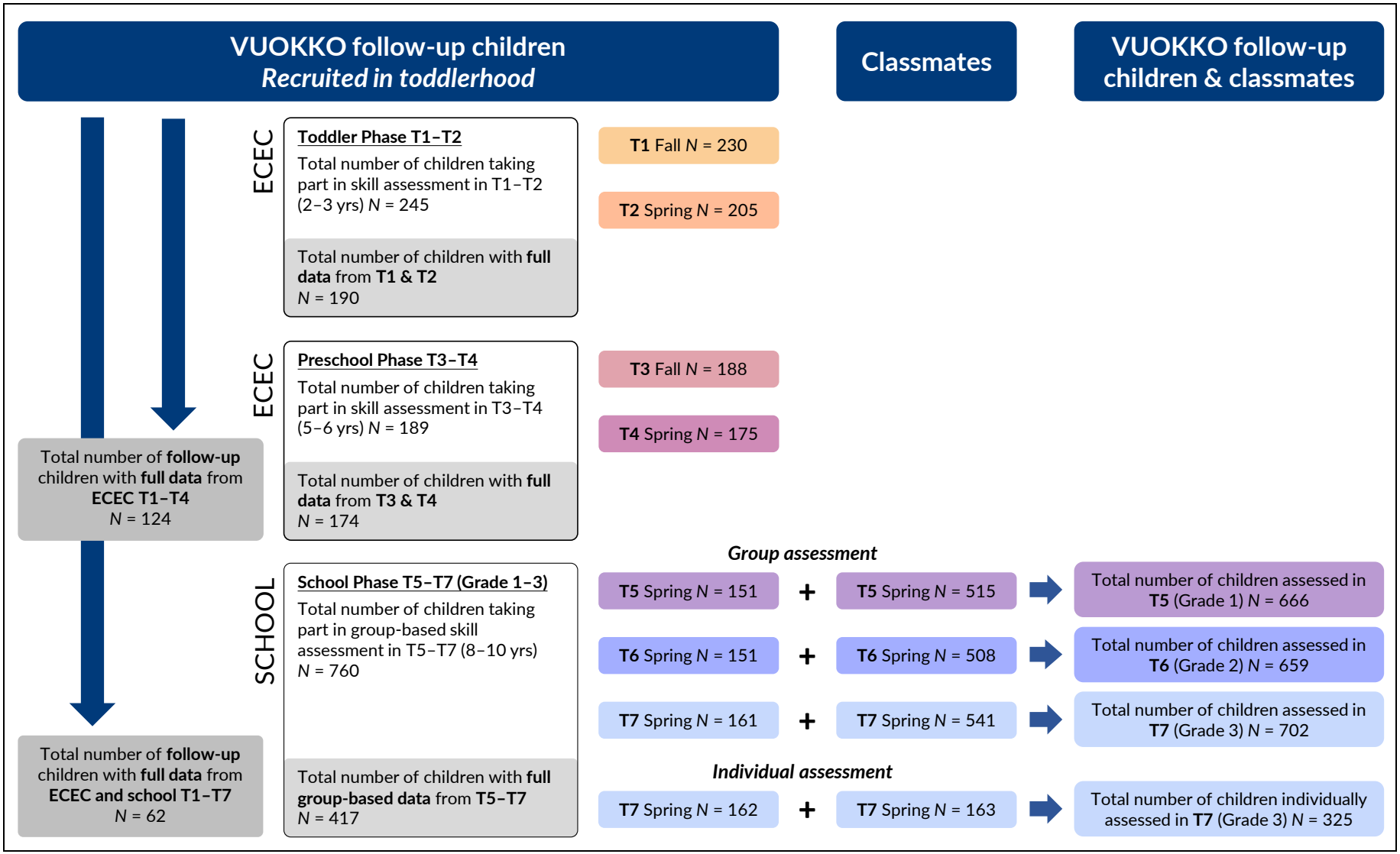
All the parents of the VUOKKO follow-up children were invited first. Out of the 122 parents who had given consent, 87 eventually agreed to participate in the assessment. In addition, 46 other biological parents of the VUOKKO follow-up children were willing to participate, meaning that altogether 132 parents of the VUOKKO follow-up children were assessed. Next, part of the parents of the children who joined the study at the school phase were invited to the assessments. Some additional exclusion criteria were used when recruiting these parents. First, the parents were not invited if their child did not participate in the assessment in Grade 1 (e.g., due to illness). In addition, the parents who had completed the Grade 1 parental questionnaire were prioritized. Finally, towards the end of the parental recruitment, after noticing that mothers, highly educated parents, and highly skilled parents had been keener to participate, in the last 30 assessments the priority was given to (1) fathers, (2) those parents who had average or below average education level and (3) those parents who had reported in the parental questionnaire that math is not easy for them.

The parents received €20 gift vouchers for participating in the study.

4.1.7. Sample Size at Each Timepoint and Attrition

Figure 2 gives an overview of the sample size at each timepoint and describes attrition at the level of participation in each assessment wave. It should be noted that in the individual tasks there may be further missingness due to assessment or participant related issues such as tiredness of the child.

FIGURE 2 An overview of the sample size at each timepoint and attrition at the level of participation in each assessment wave.



4.2. Procedures

4.2.1. CA – Children's Assessments

CA – Toddler Phase (T1–T2)

The skills of VUOKKO follow-up children were assessed first in fall 2015 and again in spring 2016. Two research assistants collected the data in the fall and three research assistants in the spring. All research assistants were students from the Faculty of Education and Psychology, and they had experience in working with young children.

All research assistants attended a three-hour training during which they were trained in the content and purpose of the young children's skill assessments. The procedure of the assessment situation, instructions on how to contact the ECEC center and the teachers of the child groups and concrete tips for the actual assessment day were carefully discussed. The ethical aspects, such as observing the children's willingness to participate and how to communicate with the teachers, were also carefully discussed (see more in Section 2). The majority of the training time was devoted to going through the individual tasks. The central aspects were how to perform them and how exactly to instruct the child, what to pay attention to in the child's behavior and how to score the children's sheets after the assessment. Research assistants practiced the completion of the assessment battery carefully with pilot children of about the same age before starting the data collection in the ECEC centers. The coordinator monitored the process via active and open discussions with the research assistants.

The contact details for the ECEC center groups and teachers were listed on a confidential paper sheet, based on which the research assistants individually contacted the teachers and set the date for completing the skill assessments in the child group. The sheet was always kept securely in a personal folder and in the possession of the research assistant while visiting the ECEC centers. At the university, the lists were securely kept in locked cabinets. After the completion of data gathering, the sheets were returned to the university and destroyed. Each research assistant was responsible for the children from an equal number of ECEC centers. Children's individual assessments in the fall were conducted between mid-November 2015 and mid-January 2016. The assessments in the spring were conducted between the end of April and mid-June 2016.

Children's individual tests were completed at the ECEC center in two separate, approximately 20-minute sections with each child. In most ECEC centers children's tasks were conducted during morning and afternoon activity hours after the naps (8:30 am to 11:00 am and 2:00 pm to 4:00 pm, respectively). For individual children the two assessment sessions were conducted on separate days to avoid putting strain on the children. The assessment took place in a separate quiet space in which as few intrusions as possible were likely to occur. Research assistants carried stuffed toy animals with them, and playfully used them to invite the children to do the tasks. In case the child felt even the least bit nervous or otherwise mildly unwilling to attend, the researcher invited a staff member from the child's group to accompany the child or postponed the assessment. The assessment space was nearly always familiar to the child and the researcher accompanied the child from and back to their own group area.

The tasks were performed in a similar way with each child. Research assistants followed a specific assessor manual which included the ways in which each task was to be explained to the child, including information on, for example, how many times the instruction was allowed to be repeated and what materials the child needed. After the assessment day, the research assistants checked and organized children's assessment sheets and scored each task according to the manual. They then returned the sheets to the research coordinator at the University. The task sheets of the participating children were stored carefully in a locked closet in a locked room. The sheets were scanned after the data gathering was complete and the data had been entered into SPSS data files and double-checked. The scans are stored on a firewall-protected JYU server.

After completing the data collection within each year, a brief meeting was held with all research assistants. Their experiences and ideas about the past data collection were carefully noted to aid in planning the subsequent data collection timepoints.

CA – Preschool Phase (T3–T4)

The skills of VUOKKO follow-up children were assessed in fall 2018 and again in spring 2019. Two research assistants collected the data in the fall and two research assistants in the spring. One of the research assistants worked with the children's assessment both in the fall and in the spring. All research assistants were students from the Faculty of Education and Psychology, and they had experience in working with young children.

All research assistants attended a three-hour training during which they were trained in the content and purpose of the preschoolers' skill assessments. The procedure of the assessment situation, instructions on how to contact the ECEC center and the teachers of the child groups and concrete tips for the actual assessment day were carefully discussed. The ethical aspects, such as observing the children's willingness to participate and how to communicate with the teachers, were also carefully discussed (see more in Section 2). The majority of the training time was devoted to going through the individual tasks. The central aspects were how to perform them and how exactly to instruct the child, what to pay attention to in the child's behavior and how to score the children's sheets after the assessment. Research assistants practiced the completion of the assessment battery carefully before starting the data collection in the ECEC centers. The coordinator monitored the process via active and open discussions with the research assistants. In the spring training, the experiences of the research assistant who had conducted the assessments in the fall were heard and she provided tips to the second research assistant.

The contact details for the ECEC center groups and teachers were listed on a confidential paper sheet, based on which the research assistants individually contacted the teachers and set the date for completing the skill assessments in the child group. The sheet was always kept securely in a personal folder and in the possession of the research assistant while visiting the ECEC centers. At the university, the lists were securely kept in locked cabinets. After the completion of data gathering, the sheets were returned to the university and destroyed. Each research assistant was responsible for the children from an equal number of ECEC centers. Children's individual assessments in the fall were conducted between the beginning of November 2018 and the end of December 2018. One child assessment was conducted in January. The assessments in the spring were conducted between the beginning of May and mid-June 2019.

Children's individual tests were completed at the ECEC center in two separate, approximately 30-minute sections with each child. In most ECEC centers children's tasks were conducted during morning and afternoon activity hours. For individual children the two assessment sessions were recommended to be conducted on separate days to avoid putting strain on the children, but in many cases children wanted to continue the tasks straight after the break. In such a case, when the child was keen to do the tasks with the researcher, all tasks were completed within one session with only short breaks for stretching and playful physical activity in between. The assessment took place in a separate quiet space in which as few intrusions as possible were likely to occur. Research assistants carried stuffed toy animals with them, and playfully used them to invite the children to do the tasks. In case the child felt even the least bit nervous or otherwise mildly unwilling to attend, the researcher invited a staff member from the child's group to accompany the child or postponed the assessment. The assessment space was nearly always familiar to the child and the researcher accompanied the child from and back to their own group area.

The tasks were performed in a similar way with each child. Research assistants followed a specific assessor manual which included the ways in which each task was to be explained to the child, including information on, for example, how many times the instruction was allowed to be repeated and what materials the child needed. After completing the tests with the child, the research assistants filled in an individual evaluation sheet to assess the child's behavior during the assessment situation.

After the assessment day, the research assistants checked and organized children's assessment sheets and scored each task according to the manual. They then returned the sheets to the research coordinator at the University. The task sheets of the participating children were stored carefully in a locked closet in a locked room. The sheets were scanned after the data gathering was complete and the data had been entered into SPSS data files and double-checked. The scans are stored on a firewall-protected JYU server.

After completing the data collection within each year, a brief meeting was held with all research assistants. Their experiences and ideas about the past data collection were carefully noted to aid in planning the subsequent data collection timepoints.

CA – School Phase: Grade 1 (T5)

The skills of the VUOKKO follow-up children and their classmates were assessed in spring 2021, when the children were in Grade 1. The children's tests were completed as group tests during one 45-minute lesson, and they were conducted in the classes during the normal school day. The assessment consisted of tasks measuring children's reading and math skills. Ten research assistants collected the data in spring 2021. All of them were either employees or students at the Faculty of Education and Psychology.

Prior to the assessments, all the research assistants attended a three-hour training during which they were trained in the content and purpose of the skill assessments. Due to COVID-19 restrictions limiting the access to research facilities at the University of Jyväskylä and the proximal contacts with other people, the training for children's skill assessment was organized as an online event. The procedure of the assessment day, instructions on how to contact the teachers and concrete tips for the actual assessment day were carefully discussed. The ethical aspects, such as observing the children's willingness to participate, how to communicate within the classroom and how to deal with consent and assessment forms were carefully discussed

(see more in Section 2). The majority of the training time was devoted to going through the individual tasks. The main aspects were how exactly to instruct the children, what to pay attention to in the child's behavior and how to score the children's sheets after the assessment. After the training, each research assistant was required to carefully practice the task instructions and assessment situation either with children or adults. The coordinator monitored the process via active and open group-discussions.

The contact details for the teachers were provided in a restricted, shared Google Drive file, where research assistants could also fill in the assessment dates and other related remarks. Research assistants individually contacted the teachers and set the date for completing the skill assessments in the classroom. The skill assessments began in the classrooms at the very beginning of April and the last assessments were conducted during the first weeks of May 2021.

While at schools, research assistants worked in pairs, so there were always two researchers in the classroom during skill assessments. One research assistant was responsible for giving the task instructions and the other was handling the papers and maintaining order in the classroom. The assessment battery included two tasks measuring children's reading skills and five tasks measuring math skills.

At the beginning of the assessment lesson, research assistants introduced themselves and explained what they were about to do with the class. The children and teachers had an opportunity to ask clarifying questions. In many classrooms the teacher was also present during the assessment. The research assistants asked the teachers not to give any additional support or advice to the individual children. Due to COVID-19 safety precautions, the research assistants maintained a safe distance from all children and teachers as well as they could and wore gloves and masks the whole time while they were in the classroom. They also disinfected the equipment such as borrowed pencils etc. after use.

The instructions for each task were performed in a similar way in every classroom. The research assistants followed a specific assessor manual which included the ways in which each task was to be explained to children, including information on, for example, how many times the instruction was allowed to be repeated, what materials children needed and what was the time limit for each task. After the assessment day, the research assistants checked and organized the children's assessment sheets and scored each task according to the manual. They then returned the sheets to the research coordinator at the University. In order to maintain fair and equal treatment for all children, in most classrooms all of the children participated in the tasks regardless of having permission to do so. For those children without permission, the sheets were taken to the university and disposed of in a secure way. None of these sheets were scored for the analysis. The task sheets of the participating children were stored carefully in a locked closet in a locked room. The sheets were scanned after the data gathering was complete and the data had been entered into SPSS data files and double-checked. The scans are stored on a firewall-protected JYU server.

After completing the data collection, a brief meeting was organized for the research assistants to share their experiences and ideas about the past data collection. The reflections were carefully noted to aid in planning the subsequent data collection timepoints.

CA – School Phase: Grade 2 (T6)

The skills of the VUOKKO follow-up children and their classmates were assessed in spring 2022, when the children were in Grade 2. The children's tests were completed as group tests during two 45-minute lessons, and they were conducted in the classes during the normal school day. The assessment consisted of tasks measuring children's reading and math skills and their interest in and anxiety about those. Eleven research assistants collected the data in spring 2022. All of them were either employees or students at the Faculty of Education and Psychology.

Prior to the assessments, all the research assistants attended a three-hour training during which they were trained in the content and purpose of the skill assessments. Due to COVID-19 restrictions limiting the access to research facilities at the University of Jyväskylä and the proximal contacts with other people, the training for children's skill assessment was organized as an online event. The procedure of the assessment day, instructions on how to contact the teachers and concrete tips for the actual assessment day were carefully discussed. The ethical aspects, such as observing the children's willingness to participate, how to communicate within the classroom and how to deal with consent and assessment forms were carefully discussed (see more in Section 2). The majority of the training time was devoted to going through the individual tasks. The main aspects were how exactly to instruct the children, what to pay attention to in the child's behavior and how to score the children's sheets after the assessment. After the training, each research assistant was required to carefully practice the task instructions and assessment situation either with children or adults. The coordinator monitored the process via active and open group-discussions.

The contact details for the teachers were provided in a restricted, shared Google Drive file, where research assistants could also fill in the assessment dates and other related remarks. Research assistants individually contacted the teachers and set the date for completing the skill assessments in the classroom. The skill assessments began in the classrooms at the very beginning of April and the last assessments were conducted during the last weeks of May 2022.

While at schools, research assistants worked in pairs, so there were always two researchers in the classroom during skill assessments. One research assistant was responsible for giving the task instructions and the other was handling the papers and maintaining order in the classroom. The assessment battery included two tasks about children's interest and anxiety in reading and math, two reading tasks, eight tasks measuring math skills and one processing speed tasks. The tasks were evenly distributed across the two lessons.

At the beginning of the first assessment lesson, research assistants introduced themselves and explained what they were about to do with the class. The children and teachers had an opportunity to ask clarifying questions. In many classrooms the teacher was also present during the assessment. The research assistants asked the teachers not to give any additional support or advice to the individual children. Due to COVID-19 safety precautions, the research assistants maintained a safe distance from all children and teachers as well as they could and wore gloves and masks the whole time they were in the classroom. However, the mask recommendation was withdrawn by the Finnish Institute for Health and Welfare while the data collection was ongoing (14 April 2022) and for approximately two-thirds of the classrooms, researchers no longer used masks and gloves while conducting the assessments in the classrooms.

The instructions for each task were performed in a similar way in every classroom. The research assistants followed a specific assessor manual which included the ways in which each task was to be explained to children, including information on, for example, how many times the

instruction was allowed to be repeated, what materials children needed and what was the time limit for each task. After the assessment day, the research assistants checked and organized the children's assessment sheets and returned them to the research coordinator at the University. All sheets and tasks were scored at the University by a group of research assistants to maintain the scoring accuracy. In order to maintain fair and equal treatment for all children, in most classrooms all of the children participated in the tasks regardless of having permission to do so. For those children without permission, the sheets were taken to the university and disposed of in a secure way. None of these sheets were scored for the analysis. The task sheets of the participating children were stored carefully in a locked closet in a locked room. The sheets were scanned after the data gathering was complete and the data had been entered into SPSS data files and double-checked. The scans are stored on a firewall-protected JYU server.

After completing the data collection, a brief meeting was organized for the research assistants to share their experiences and ideas about the past data collection. The reflections were carefully noted to aid in planning the subsequent data collection timepoints.

CA – School Phase: Grade 3, Group Assessment (T7)

The skills of the VUOKKO follow-up children and their classmates were assessed in spring 2023, when the children were in Grade 3. The children's tests were completed as group tests during two 45-minute lessons, and they were conducted in the classes during the normal school day. The assessment consisted of tasks measuring children's reading and math skills and their interest in and anxiety about those. Twelve research assistants collected the data in spring 2023. All of them were either employees or students at the Faculty of Education and Psychology.

Prior to the assessments and training, the tasks were piloted in one classroom by one research assistant. All the research assistants attended a three-and-half-hour training during which they were trained in the content and purpose of the skill assessments. The procedure of the assessment day, instructions on how to contact the teachers and concrete tips for the actual assessment day were carefully discussed. The ethical aspects, such as observing the children's willingness to participate, how to communicate within the classroom and how to deal with consent and assessment forms were carefully discussed (see more in Section 2). The majority of the training time was devoted to going through the individual tasks. The main aspects were how exactly to instruct the children, what to pay attention to in the child's behavior and how to score the children's sheets after the assessment. After the training, each research assistant was required to carefully practice the task instructions and assessment situation either with children or adults. Two weeks after the first training, another meeting was held with the research assistants as an online meeting. The purpose of the meeting was to go through the scoring principles and to share thoughts and experiences about practicing the tasks, and to ask questions about the instructions, tasks and the upcoming assessment situations in general. The coordinator monitored the process via active and open group-discussions.

The contact details for the teachers were provided in a restricted, shared OneDrive file, where research assistants could also fill in the assessment dates and other related remarks. Research assistants individually contacted the teachers and set the date for completing the skill assessments in the classroom. The skill assessments began in the classrooms at the beginning of March and the last assessments were conducted after mid-May 2023.

While at schools, research assistants worked in pairs, so there were always two researchers in the classroom during skill assessments. One research assistant was responsible for giving the

task instructions and the other was handling the papers and maintaining order in the classroom. The assessment battery included two tasks about children's interest and anxiety in reading and math, two reading tasks and nine tasks measuring math skills. The tasks were evenly distributed across the two lessons.

At the beginning of the first assessment lesson, research assistants introduced themselves and explained what they were about to do with the class. The children and teachers had an opportunity to ask clarifying questions. In many classrooms the teacher was also present during the assessment. The research assistants asked the teachers not to give any additional support or advice to the individual children.

The instructions for each task were performed in a similar way in every classroom. The research assistants followed a specific assessor manual which included the ways in which each task was to be explained to children, including information on, for example, how many times the instruction was allowed to be repeated, what materials children needed and what was the time limit for each task. After the assessment day, the research assistants checked and organized the children's assessment sheets and returned them to the research coordinator at the University. All sheets and tasks were scored at the University by a group of research assistants to maintain the scoring accuracy. In order to maintain fair and equal treatment for all children, in most classrooms all of the children participated in the tasks regardless of having permission to do so. For those children without permission, the sheets were taken to the university and disposed of in a secure way. None of these sheets were scored for the analysis. The task sheets of the participating children were stored carefully in a locked closet in a locked room. The sheets were scanned after the data gathering was complete and the data had been entered into SPSS data files and double-checked. The scans are stored on a firewall-protected JYU server.

After completing the data collection, the research assistants were asked to give feedback about their experiences and ideas about the past data collection by filling in a short feedback form. They were asked to share their thoughts about, for example, the training, the assessment situations and data collection timetable, the tasks concerning scoring and in general about their feelings towards carrying out the data collection.

CA – School Phase: Grade 3, Individual Assessment (T7)

The skills of the VUOKKO follow-up children and some of their classmates were assessed with individual tests in spring 2023, when the children were in Grade 3. One assessment took about 30 minutes, and they were conducted during the normal school day in a quiet and calm room in the school. The assessment consisted of tasks measuring children's reading and math skills. Six research assistants collected the data in spring 2023. All of them were students from the Faculty of Education and Psychology.

Prior to the assessments and training, the tasks were piloted with a few children in two different classrooms by one research assistant. All the research assistants attended a ten-hour training on three different days. During the training they were trained in the content and purpose of the skill assessments. The procedure of the assessment day, instructions on how to contact the teachers and concrete tips for the actual assessment were carefully discussed. The ethical aspects, such as observing the children's willingness to participate and how to deal with consent and assessment sheets were carefully discussed (see more in Section 2). The majority of the training time was devoted to going through the individual tasks: how exactly to instruct the children and what to pay attention to. After the training, each research assistant was required

to carefully practice the task instructions and assessment situation with children or adults. In addition, they piloted each other trying to keep the situation as authentic as possible. The coordinator monitored the process via one-on-one discussions, group discussions, and checking a couple of assessment audio recordings and test scorings in the test papers per research assistant.

The contact details for the teachers were provided in a restricted, shared OneDrive file, where research assistants could also fill in the assessment dates and other related remarks. From that file they also could see the dates set for the group assessment and observation to avoid visiting the class on the same day. Research assistants took care of contacting the teachers either by e-mail or by phone to arrange the days for the assessments. Teachers were told the amount of the children that would be invited to participate from their class. In addition, the timetable was described to the teachers and all the possible practical matters were discussed. Children's individual assessments started at the beginning of March 2023, and the last assessment was completed at the beginning of May 2023.

The instructions for each task were performed in a similar way in each assessment. The research assistants followed a specific assessor manual which included the ways in which each task was to be explained to children, including information on, for example, how many times the instruction was allowed to be repeated, what materials children needed and what was the possible time limit for each task. After the individual assessment session, the research assistants filled in an individual evaluation sheet to assess the child's behavior during the assessment situation. Voice recordings were gathered in the assessment situation and the recordings were used to support scoring of the individual skill assessments.

After the assessment day, the research assistants checked and organized children's assessment sheets and returned them to the University. All sheets and tasks were scored at the University by the research assistants to maintain the scoring accuracy. The task sheets of the participating children were stored carefully in a locked closet in a locked room. The sheets were scanned after the data gathering was complete and the data had been entered into SPSS data files and double-checked. The scans are stored on a firewall-protected JYU server.

After completing the data collection, a brief meeting was organized for the research assistants to share their experiences and ideas about the past data collection. The reflections were carefully noted to aid in planning the subsequent data collection timepoints.

4.2.2. CR – Children's Reports

CR – Preschool Phase (T3–T4)

In fall 2018 and in spring 2019, at the end of the child's individual skill assessment, the child was interviewed on their interest in tasks related to letters and reading as well as to numbers and math both at home and in ECEC. The research assistant gave the child instructions on how to answer the questions, using a scale of five faces. Each item was read aloud by the research assistant and the child was asked to point to the option that best suited them. All the children who participated in the individual assessment were also interviewed about their task interest. This means that the data are available from 188 children in the fall and 175 in the spring.

CR – School Phase: Grade 2 (T6)

At the beginning of the children's group assessment, they were asked to fill in two short paper questionnaires on their interest in reading and math and reading and math anxiety. The research assistant gave the children instructions on how to fill in the questionnaires and explained the meanings of the scales. Each item was read aloud one by one by the research assistant and the children were asked to cross out the option that best suited them. All the children who participated in the group assessment also filled in the questionnaires, meaning questionnaire data are available from 659 children.

CR – School Phase: Grade 3 (T7)

At the beginning of the children's group assessment, they were asked to fill in two short paper questionnaires on their interest in reading and math and reading and math anxiety. The research assistant gave the children instructions on how to fill in the questionnaires and explained the meanings of the scales. Each item was read aloud one by one by the research assistant and the children were asked to cross out the option that best suited them. All the children who participated in the group assessment also filled in the questionnaires, meaning questionnaire data are available from 702 children.

4.2.3. PA – Parental Assessments

The parents of the participating children were invited to the individual assessment of their own reading and mathematics skills. One assessment took circa 90 minutes, of which the tasks took approximately 75 minutes, and the rest of the time was used for possible questions and giving feedback. The parental assessments took place on the premises of JYU. The first assessment was conducted in June 2021, and the last assessment was completed in November 2022. Nine research assistants collected the parental skill data. All of them were employees or students at the Faculty of Education and Psychology.

The research assistants were carefully trained for the data collection – the training and piloting took altogether approximately 15 hours. In the training sessions the research assistants were introduced to the research topic, the ethical principles, and conducting the assessments. Each task was carefully practiced to make sure all the research assistants would give the instructions to the parents in the same way. The research assistants practiced the assessment situation by piloting each other and other adults, trying to keep the situation as authentic as possible. The coordinator monitored the process via one-on-one discussions, group-discussions, and checking a couple of assessment audio recordings and test scorings in the test papers per research assistant.

Six research assistants took care of contacting the parents to invite them to the assessments. When making the calls, they could see other research assistants' timetables from a common Google calendar. The assessment bookings were added to the shared calendar without parents' information. The more detailed information of each assessment was shared via the firewall-protected JYU server using an encrypted Excel file. The protected Excel file included information needed in the assessment situation, such as parents' names and contact details.

The assessment battery consisted of 12 tasks, including four reading tasks and eight tasks measuring mathematics skills. The parents were told it is possible to have a short break in

between the tasks, if needed. During the last task (KTLT-A, 30 minutes), the researcher exited the room to let the parent complete the task alone in peace. The researcher had a Zoom connection to the room so that they could see and hear the parent all the time. During the last task the research assistants followed the screen and scored the tasks, prioritizing the tasks about which the parents would get the feedback if they wanted to have it. Feedback was offered for the reading tasks (*Word list, Pseudoword list, RAN* and *Text reading*) and four mathematics tasks (*Multiplication, Arithmetic, RMAT* and *KTLT-A*) of which there was adult norm data available. After the assessment situation the rest of the tasks were scored. Voice recordings were gathered from the reading tasks (see above) and the recordings were used to support scoring of those tasks.

The instructions for each task were performed in a similar way in every assessment. Research assistants followed a specific assessor manual which included the ways in which each task was to be explained to parents, including information on, for example, how the tasks were to be completed, what materials parents needed and what was the possible time limit for each task. After the assessment, the research assistants finished the scoring if needed and organized the parent's sheets. The task sheets were stored carefully in a locked closet in a locked room. The sheets were scanned after the data gathering was complete and the data had been entered into SPSS data files and double-checked. The scans are stored on a firewall-protected JYU server.

Activities during the COVID-19 period followed the rules of the Faculty of Education and Psychology. Prior to the assessments, the research assistants disinfected the tables, chairs and door handles. The parents were given the instruction to arrive to the assessment situation only if they felt fully healthy, and they were asked to inform the researcher and postpone the assessment if they got sick, if they were having symptomless COVID-19 or if they knew they had been potentially exposed to the coronavirus. During the assessment both the parent and the researcher used face masks (FFP2 level). The exception to this was during the last task (KTLT-A) when the parent was alone in the room – they were allowed to be there without a mask. Besides the masks the researchers also wore disposable gloves and maintained a safe distance between them and the parent. After the assessment the researchers disinfected the table, chairs and door handles again as well as the pens and laminated task sheets.

During the data collection, a few meetings were organized for the research assistants to share their experiences and ideas about the training and data collection in general. The reflections were carefully noted to aid in planning the subsequent data collection.

4.2.4. PQ – Parental Questionnaires

PQ – Toddler Phase (T1–T2)

Parental questionnaires with an introduction letter attached were mainly sent to parents' home addresses. The questionnaires were sent to the parent who had filled in the consent form. This led to receiving one questionnaire per child and the respondent could be either the mother or the father. In cases where parents did not include their home address on the informed consent form, the questionnaire was delivered to parents via the ECEC centers and teachers. Parental questionnaires were sent out in October 2015. A polite reminder and a new questionnaire was sent to non-responding parents at the beginning of February 2016. Altogether 245 questionnaires were sent out and parents of 208 children (85%) returned the questionnaire, of which 192 were mothers and 16 fathers.

PO – Preschool Phase (T3–T4)

Parental questionnaires with an introduction letter attached were mainly sent to parents' home addresses. In cases where parents did not include their home address on the informed consent form, the questionnaire was delivered to parents via the ECEC centers and teachers. Parental questionnaires were sent out in October 2018. A polite reminder and a new questionnaire was sent to non-responding parents at the end of January 2019. Altogether 168 questionnaires were sent out and parents of 134 children (80%) returned the questionnaire, of which 123 were mothers and 9 fathers, and in two cases both parents had been marked as respondents.

PO – School Phase: Grade 1 (T5)

Parental questionnaires with an introduction letter attached were sent to parents' home addresses in early June 2021. Due to the summer holiday season, a polite reminder with a new questionnaire was sent to non-responding parents at the end of September 2021. Altogether 596 questionnaires were sent out and parents of 354 children (59%) returned the questionnaire. In addition, 101 children's other biological parents (see section 4.1.6.) were invited to fill in the questionnaire either at the university after the parental assessment or at home, and 92 of them finally returned the questionnaire. Eventually the parental questionnaire data from Grade 1 is available from parents of 446 children. Of those parents, 300 were mothers, 143 were fathers, and in two cases the respondent was someone else and in one questionnaire the information about the respondent was not available. Of those 446 responses, 393 parents replied to questionnaires concerning both their child and themselves, and 53 parents replied to questionnaires concerning only themselves. This is because some of those parents, who were asked to fill in the parental questionnaire while they were invited to participate in the parental assessment, filled in the questionnaire when their child had already been in Grade 2 for a few months. In these cases, the parents were not asked to fill in the questionnaires concerning their child in Grade 1 anymore. From 83 families both parents filled in the questionnaire. The parents had an opportunity to reply to either the paper questionnaire or online survey (Webropol). They were offered a link to the online survey in the letter they were sent with the paper questionnaire. Eventually parents of 115 children replied to the online survey, and parents of 331 children replied to the paper questionnaire.

PO – School Phase: Grade 2 (T6)

Parental questionnaires with an introduction letter attached were sent to parents' home addresses in early May 2022. The questionnaire was sent to only one parent of each participating family. A polite reminder with a new questionnaire was sent to non-responding parents in the middle of June 2022 and again in July. Altogether 684 questionnaires were sent out and parents of 370 children (54%) returned the questionnaire. Of those 370 respondents, 291 were mothers, 74 were fathers, and in two cases the respondent was someone else and in one questionnaire the information about the respondent was not available. Besides, in two cases both parents had been marked as respondents. The parents had an opportunity to reply to either the paper questionnaire or online survey (Webropol). They were offered a link to the online survey in the letter they were sent with the paper questionnaire. Eventually parents of 106 children replied to the online survey, and parents of 264 children replied to the paper questionnaire.

PO – School Phase: Grade 3 (T7)

Parental questionnaires with an introduction letter attached were sent as an online survey (Webropol) to parents' email addresses after mid-May 2023. In the cases where parents did not include their email address on the informed consent form, or if the email address did not work, the paper questionnaire was sent to their home addresses. In addition, the parents who received the online survey were offered an opportunity to receive the paper questionnaire by mail by contacting the research group. A polite reminder with a link to the online survey was sent to non-responding parents' emails at the beginning of June. Another reminder was sent as a paper questionnaire to non-responding parents' home addresses towards the end of June.

Altogether 954 questionnaires were sent out and parents of 493 children (52%) returned the questionnaire. Of those 493 respondents, 362 were mothers, 126 were fathers, and in two cases the respondent was someone else and in three questionnaires the information about the respondent was not available. The questionnaire was sent to all of those parents who had given consent to fill in the parental questionnaire, meaning that the questionnaires were sent to both of a child's parents in 152 families. From 55 families both parents filled in the questionnaire. Eventually parents of 381 children replied to the online survey, and parents of 112 children replied to the paper questionnaire.

Two different versions of the questionnaires were used: the parents, who had already filled in the parental questionnaire when their child was either in Grade 1 or in the Grade 2 were sent a shorter version of the questionnaire. Whereas the parents, who had not filled in the parental questionnaire in those years, were sent a longer version of the questionnaire. The longer version included some questions that the above-mentioned parents had already replied to in the previous years. Of the returned questionnaires, 381 were the short versions and 112 were the long versions.

4.2.5. TCI – Teacher–Child Interaction Quality (Classroom Observations)

TCI – Toddler Phase (T1–T2)

The assessment of the teacher–child interaction was based on video recordings obtained from toddler classrooms. Observations were conducted on two different days in 32 ECEC toddler classrooms and during one day in 10 classrooms. Video recording was completed in 41 classrooms. A live observation using the CLASS Toddler observation instrument (see section 5.6.1 for more details) was conducted in one classroom, in which teachers were not willing to be video recorded. Live observation allowed information on the quality of the teacher–child interaction to be obtained under circumstances when video recording was not a possibility, and it was performed by the same CLASS-certified researcher in every classroom. Observations began at the end of January 2016 and the last observations were conducted by the end of May 2016. One video camera (Canon Legria) was used in every classroom to capture the interaction between the teacher and the children.

Observations were usually conducted between 8 a.m. and 12 a.m., during the time between breakfast and lunch. Video recordings were conducted mainly by one and the same research assistant in all ECEC centers, and four observation visits were conducted by two research assistants. All the research assistants were students from the Faculty of Education and Psychology. During observation days, information about the classroom space was gathered

along with details on the number of children and teachers present during the observation and the content of the activity being video recorded. Prior to starting the video recording, the observer spent some time with the children and discussed with the teachers in the classroom. Children had an opportunity to see themselves from the video camera and familiarize themselves with the observer. In the observation situation, while the video recording was on, the researcher aimed to engage in interaction as little as possible to avoid interfering with the natural flow of events. Children and teachers were given a possibility to stop the recording at any moment if they felt uncomfortable in the situation. Only those children and teachers who had a consent to be video recorded were allowed to be seen on the recording. Children without consent were either with other teachers in another room or situated out of the camera frame. The decision was made situationally in collaboration with the teacher and careful notes were kept by the researcher.

The aim was to video record at least four videos in every classroom, representing the four activities presented below. The anticipated length of each activity, and thus, of each video recording, was between 10 and 20 minutes. The focus of the observation was on typical daily activities during which the children and at least one teacher were engaging in the same activity. Since the focus of the observation was on the interaction between teachers and children, the teacher had to be present and observable on the video for at least a third of the video-recorded time. The four activities recorded represented the typical daily experiences for children in ECEC (see, e.g., Booren et al., 2012; Slot et al., 2016): *play* (i.e., joint play with teacher and children); *care routines* (i.e., mealtime, teacher present at least in the monitoring role); *emerging academic activities* (e.g., pre-literacy or pre-numeracy skills during a group activity or circle time (whichever was more appropriate/typical for the classroom); *creative activities* (e.g., craft, music, movement and/or dance). Altogether 154 videos were recorded across the 41 classrooms. In five classrooms, only three activities were recorded due to daily schedules not meeting the requirements for the observation protocol. In the remaining classrooms all four activities were observed. The length of the individual video recordings varied from 10 minutes to 25 minutes (average 17 minutes). The total duration of video-recorded activities per toddler classroom ranged from 44 minutes to 82 minutes (average 67 minutes per classroom).

TCI – Preschool Phase (T3–T4)

The assessment of the teacher–child interaction was based on video recordings obtained from preschool classrooms. Observations were conducted during one day in each of the participating 56 preschool classrooms. Video recording was completed in 51 classrooms. A live observation using the CLASS Pre-K observation instrument (see section 5.6.3. for more details) was conducted in five classrooms, in which teachers were not willing to be video-recorded. Live observation allowed information on the quality of the teacher–child interaction to be obtained under circumstances when video recording was not a possibility, and it was performed by the same CLASS-certified researcher in every classroom. Observations began at the beginning of February 2019 and the last observations were conducted by the end of May 2019. One video camera (Canon Legria) was used in every classroom to capture the interaction between the teacher and the children.

Observations were conducted between 8 a.m. and 12 a.m., during the time between breakfast and lunch. Video recordings were conducted by one and the same researcher in all ECEC centers. During observation days, information about the classroom space was gathered along with details on the number of children and teachers present during the observation and the content of the activity being video recorded. Prior to starting the video recording, the observer

spent some time with the children and discussed with the teachers in the classroom. Children had an opportunity to see themselves from the video camera and familiarize themselves with the observer. In the observation situation, while the video recording was on, the researcher aimed to engage in interaction as little as possible to avoid interfering with the natural flow of events. Children and teachers were given a possibility to stop the recording at any moment if they felt uncomfortable in the situation. Only those children and teachers who had consent to be video recorded were allowed to be seen on the recording. Children without consent were either with other teachers in another room or situated out of the camera frame. The decision was made situationally in collaboration with the teacher and careful notes were kept by the researcher.

At least four videos were recorded in every classroom. The focus of the observation was on typical daily activities during which the children and at least one teacher were engaging in the same activity. Since the focus of the observation was on the interaction between teachers and children, the teacher had to be present and observable on the video for at least a third of the video recorded time. The four activities recorded represented the typical daily experiences for children in ECEC (see, e.g., Booren et al., 2012; Slot et al., 2016): *play* (i.e., joint play with teacher and children); *care routines* (i.e., mealtime, teacher present at least in the monitoring role); *emerging academic activities* (e.g., pre-literacy or pre-numeracy skills during a group activity or circle time (whichever was more appropriate/typical for the classroom); *creative activities* (e.g., craft, music, movement and/or dance). Altogether 187 videos were recorded across the participating classrooms (excluding the five live-observed classrooms). In 22 classrooms all four activities were captured, in 25 classrooms three activities were captured and in four classrooms two activities were captured. Across the classrooms, creative activity was the most often missed activity, followed by lunch and play. The length of the individual video recordings varied from 10 minutes to 56 minutes (average 16 minutes). The total duration of video recorded activities per preschool classroom ranged from 47 minutes to 137 minutes (average 82 minutes per classroom).

TCI – School Phase: Grade 1 (T5)

Observations were conducted on one day in each of the 37 participating Grade 1 classrooms. Researchers recorded on video one math lesson and one lesson of mother tongue/literacy with two video cameras: One was placed at the back of the classroom (Canon Legria) and another at the front (GoPro) to capture both the teacher view (facing the students) and the student view (facing the teacher) of the classroom. In addition, teachers wore a collar microphone to better capture their voices. This was deemed necessary because due to COVID-19 pandemic, teachers wore a mask or visor during lessons. Only those children and teachers who had consent to be video recorded were allowed to be seen on the recording. Children without consent were either with another teacher in another space or other part of the classroom, situated out of the camera frame, or present but blurred afterwards from the recordings. The decision was made situationally in collaboration with the teacher and careful notes were kept by the research assistants.

Video recordings were conducted by five research assistants, who were carefully trained. All of them were students from the Faculty of Education and Psychology. Because the COVID-19 pandemic limited access to research facilities at the University of Jyväskylä and proximal contacts with other people, the training of research assistants for classroom observations was organized online. Research assistants first attended a two-hour training during which they were trained in the content and purpose of the observation. The procedure of the observation day,

instructions on how to contact the teachers and concrete tips for fieldwork were carefully discussed. Furthermore, the training included going through the ethical issues, including data protection, confidentiality, consent procedures and incidental findings protocol (see more in Section 2). A couple of days later, the research assistants attended another one-hour online session, during which the technical details of audio- and video-recording equipment were explained and concretely demonstrated. Each observer was instructed to practice the use of the cameras and audio recorder before going to schools.

The contact details for the teachers were provided in a restricted, shared Google Drive file, where research assistants could also fill in the observation dates and other related remarks. Research assistants individually contacted the teachers and set the date for observation in the classroom once the access to schools was again allowed after the COVID-19 closure. Although researchers had access to classrooms, COVID-19 nevertheless impacted the observations in the classrooms. For instance, some classrooms were quarantined for several days when a positive COVID-19 infection in the classroom was identified. Observations were conducted in the spring term of Grade 1, from the end of February until the end of April 2021.

One research assistant was present in the classroom during the observation. At the beginning of the lesson, they introduced themselves and explained what the observation was about. Children and teachers had an opportunity to ask clarifying questions. During the recording, research assistants in most cases positioned themselves toward the back or sides of the classroom space and aimed at being as invisible as possible to avoid interfering with the regular classroom practices. Due to COVID-19 safety precautions, they also maintained a safe distance from all children and teachers and wore gloves and masks the whole time they were in the classroom. They also disinfected the equipment after use.

In addition to video recording the two lessons, research assistants also asked teachers an interview question about the individual characteristics of children in their classroom which might impact the way teachers interact with the children in typical classroom situations. Such characteristics could relate to learning, behavior, social-emotional skills or L2 status, just to name few. This information was recorded as a written field-note and later inserted into an independent SPSS file with ID numbers and classified codes.

Altogether 36 math lessons and 34 mother tongue lessons were recorded. In two classrooms it was not possible to record video of mother tongue lessons due to mixed grouping of parallel classes during these lessons (i.e., children without consent for recording being present). In these cases, environmental studies were recorded instead. In one classroom both math and mother tongue lessons were replaced with environmental studies lessons. The length of the video recordings varied from 19 minutes to 70 minutes (average 41 minutes). The total duration of activities recorded on video per Grade 1 classroom ranged from 58 minutes to 120 minutes (average 83 minutes per classroom).

TCI – School Phase: Grade 2 (T6)

Observations were conducted on one day in each of the 36 participating Grade 2 classrooms. Researchers recorded on video one math lesson and one lesson of mother tongue/literacy with two video cameras: One was placed at the back of the classroom (Canon Legria) and another at the front (GoPro) to capture both the teacher view (facing the students) and the student view (facing the teacher) of the classroom. In addition, teachers wore a collar microphone to better capture their voices. This was deemed necessary because due to COVID-19 pandemic, teachers

wore a mask or visor during lessons. Only those children and teachers who had consent to be video recorded were allowed to be seen on the recording. Children without consent were either with another teacher in another space or other part of the classroom, situated out of the camera frame or present but blurred afterwards from the recordings. The decision was made situationally in collaboration with the teacher and careful notes were kept by the research assistants.

Video recordings were conducted by six research assistants, who were carefully trained. All of them were students from the Faculty of Education and Psychology. Because the COVID-19 pandemic still limited access to research facilities at the University of Jyväskylä and proximal contacts with other people, the training of research assistants for classroom observations was organized online. Research assistants attended a two-hour training during which they were trained in the content and purpose of the observation. The procedure of the observation day, instructions on how to contact the teachers and concrete tips for fieldwork were carefully discussed. Furthermore, the training included going through the ethical issues, including data protection, confidentiality, consent procedures and incidental findings protocol (see more in Section 2). A couple of days later, the research assistants attended a more practical training which was organized in live mode with the permission from the University and with only six people present and wearing masks. In this one-hour session the technical details of audio- and video-recording equipment were explained, and research assistants had an opportunity to concretely try out the devices. Each observer was instructed to practice the use of the cameras and audio recorder before going to schools.

The contact details for the teachers were provided in a restricted, shared Google Drive file, where research assistants could also fill in the observation dates and other related remarks. Research assistants individually contacted the teachers and set the date for observation in the classroom once the access to schools was again allowed after the COVID-19 closure. Observations were conducted in the spring term of Grade 2, from mid-March until the end of May 2022.

One research assistant was present in the classroom during the observation. At the beginning of the lesson, they introduced themselves and explained what the observation was about. Children and teachers had an opportunity to ask clarifying questions. During the recording, research assistants in most cases positioned themselves toward the back or sides of the classroom space and aimed at being as invisible as possible to avoid interfering with the regular classroom practices. Due to COVID-19 safety precautions, at the beginning of the observations, they also maintained a safe distance from all children and teachers and wore gloves and masks the whole time they were in the classroom. However, the mask recommendation was withdrawn by the Finnish Institute for Health and Welfare while the data collection was ongoing (14 April 2022) and for approximately two-thirds of the classrooms, researchers no longer used masks and gloves while conducting the observations.

In addition to video recording the two lessons, research assistants also asked teachers an interview question about the individual characteristics of children in their classroom, which might impact the way teachers interact with the children in typical classroom situations. Such characteristics could relate to learning, behavior, social-emotional skills or L2 status, just to name few. This information was recorded as a written field-note and later inserted into an independent SPSS file with ID numbers and classified codes.

Altogether 35 math and mother tongue lessons were recorded across the participating classrooms. For two parallel classrooms, observation was conducted by combining all children

with consent for video recording from these classrooms into one group. The classroom teachers both taught this newly formed group of children, one during a math lesson and the other during a mother tongue lesson. The length of the video recordings varied from 21 minutes to 89 minutes (average 42 minutes). The total duration of activities recorded on video per Grade 2 classroom ranged from 57 minutes to 122 minutes (average 80 minutes per classroom).

TCI – School Phase: Grade 3 (T7)

Observations were conducted on one day in each of the 39 participating Grade 3 classrooms. Researchers recorded on video one math lesson and one lesson of mother tongue/literacy with two video cameras: One was placed at the back of the classroom (Canon Legria) and another at the front (GoPro) to capture both the teacher view (facing the students) and the student view (facing the teacher) of the classroom. In addition, teachers wore a collar microphone to better capture their voices. Only those children and teachers who had consent to be video recorded were allowed to be seen on the recording. Children without consent were either with another teacher in another space or other part of the classroom, situated out of the camera frame or present but blurred afterwards from the recordings. The decision was made situationally in collaboration with the teacher and careful notes were kept by the research assistants.

Video recordings were conducted by eight research assistants, who were carefully trained. All of them were students from the Faculty of Education and Psychology. Research assistants attended a three-hour live training during which they were trained in the content and purpose of the observation. The first hour of training concerned the procedure of the observation day, instructions on how to contact the teachers and concrete tips for fieldwork. Furthermore, the training included going through the ethical issues, including data protection, confidentiality, consent procedures and incidental findings protocol (see more in Section 2). The second part of the training concerned the technical details of audio and video recording equipment along with discussing the placement of the cameras in the classrooms as well as with their purposes. Research assistants had an opportunity to concretely try out the devices. Each observer was instructed to practice the use of the cameras and audio recorder before going to schools.

The contact details of the teachers were provided in a restricted, shared OneDrive file, where research assistants could also fill in the observation dates and other related remarks and simultaneously see on which day child group and individual assessments were to be conducted in the classrooms. Research assistants individually contacted the teachers and set the date for observation in the classroom. Observations were conducted in the spring term of Grade 3, from the beginning of March until the end of May 2023.

One research assistant was present in the classroom during the observation. At the beginning of the lesson, they introduced themselves and explained what the observation was about. Children and teachers had an opportunity to ask clarifying questions. During the recording, research assistants in most cases positioned themselves toward the back or sides of the classroom space and aimed at being as invisible as possible to avoid interfering with the regular classroom practices.

Altogether 39 math and 38 mother tongue lessons were recorded across the participating classrooms. The length of the video recordings varied from 23 minutes to 54 minutes (average 38 minutes). The total duration of activities recorded on video per Grade 3 classroom ranged from 61 minutes to 93 minutes (average 77 minutes per classroom).

4.2.6. TQ – Teachers' Questionnaires

TO – Toddler Phase (T1–T2)

Teachers' questionnaires with an introduction letter attached were sent by mail to each ECEC classroom in February 2016. The teachers were informed about the upcoming questionnaire by email, approximately one week before the questionnaires arrived. One teacher questionnaire was sent to each participating teacher in each ECEC classroom according to their consent. We encouraged all teachers, though at least one from each classroom, to return the questionnaire. In case the teachers did not return their questionnaire, a polite reminder was sent via email, and a new questionnaire was mailed if needed. Altogether 145 questionnaires were sent and 117 were returned (81%). In the end at least one teacher from each participating classroom ended up returning the teacher's questionnaire. More than one teacher had completed the questionnaire in 37 classrooms.

TO – Preschool Phase (T3–T4)

Teachers' questionnaires with an introduction letter attached were sent by mail to each ECEC classroom at the beginning of March 2019. The teachers were informed about the upcoming questionnaire by email, approximately one week before the questionnaires arrived. One teacher questionnaire was sent to each participating teacher in each ECEC classroom according to their consent. We encouraged all teachers, though at least one from each classroom, to return the questionnaire. In case the teachers did not return their questionnaire, a polite reminder was sent via email, and a new questionnaire was mailed if needed. Altogether 152 questionnaires were sent and 98 were returned (64%). In the end at least one teacher from 50 (out of 57) classrooms ended up returning the teacher's questionnaire. More than one teacher had completed the questionnaire in 31 classrooms.

TO – School Phase: Grade 1 (T5)

Teachers' questionnaires, with an introduction letter attached, were sent by mail to schools at the end of February 2021. The teachers were informed about the upcoming questionnaire by email, approximately one week before the questionnaires arrived. In case the teachers did not return their questionnaire, a polite reminder was sent via email, and a new questionnaire was mailed if needed. Of the 42 participating teachers, 40 returned the questionnaire (95%).

TO – School Phase: Grade 2 (T6)

Teachers' questionnaires, with an introduction letter attached, were sent by mail to schools at the end of March 2022. The teachers were informed about the upcoming questionnaire by research assistants, while they were visiting the classroom for observations or child assessments. In case the teachers did not return their questionnaire, a polite reminder was sent via email, and a new questionnaire was mailed if needed. Of the 41 participating teachers, 40 teachers (98%) returned the questionnaire.

TO – School Phase: Grade 3 (T7)

Teachers' questionnaires, with an introduction letter attached, were sent by mail to schools at the beginning of May 2023. The teachers were informed about the upcoming questionnaire by

research assistants while they were visiting the classroom for observations or child assessments. In case the teachers did not return their questionnaire, a polite reminder was sent via email, and a new questionnaire was mailed if needed. Of the 54 participating teachers, 45 teachers (83%) returned the questionnaire.

4.2.7. TR – Teachers' Reports Concerning Individual Children

TR – Toddler Phase (T1-T2)

Teachers' reports on individual children, with an introduction letter attached, were sent out to each ECEC classroom twice during the year: in October 2015 and in May 2016. Teachers evaluated the same follow-up children on both timepoints. The teachers were informed about the upcoming reports by email, approximately one week before the reports arrived. One report on individual children (that included all participating children) was sent to each participating teacher in each ECEC classroom according to their own consent. We encouraged all teachers, but at least one from each classroom, to return the report on individual children.

In fall 2015, at least one teacher returned the report from each participating classroom. Altogether 136 reports were sent and 94 were returned (69%). Individual reports were completed for 228 children.

In spring 2016, at least one teacher returned the report from 35 (out of 43) participating classrooms. Altogether 146 reports were sent and 96 were returned (66%). Individual reports were completed for 192 children.

TR – Preschool Phase (T3–T4)

Teachers' reports on individual children, with an introduction letter attached, were sent out to each ECEC classroom twice during the year: in November 2018 and in May 2019. Teachers evaluated the same follow-up children on both timepoints. The teachers were informed about the upcoming reports by email, approximately one week before the reports arrived. One report on individual children (that included all participating children) was sent to each participating teacher in each ECEC classroom according to their own consent. We encouraged all teachers, but at least one from each classroom, to return the report on individual children.

In fall 2018, at least one teacher returned the report from 55 (out of 57) participating classrooms. Altogether 152 reports were sent and 106 were returned (70%). Individual reports were completed for 162 children.

In spring 2019, at least one teacher returned the report from 49 (out of 57) participating classrooms. Altogether 152 reports were sent and 94 were returned (62%). Individual reports were completed for 136 children.

TR – School Phase: Grade 1 (T5)

Teachers' reports on individual children, with an introduction letter attached, were sent out to each classroom at the beginning of April 2021. Teachers were informed about the upcoming reports by email, approximately one week before the reports arrived. Teachers only completed

the reports on individual children of the VUOKKO follow-up children in their classroom (1–8 children/classroom). Out of those 42 teachers who had initially agreed to complete the report on individual children, 41 returned it (98%). Individual reports were completed for 129 children.

TR – School Phase: Grade 2 (T6)

Teachers' reports on individual children, with an introduction letter attached, were sent out to each classroom in mid-April 2022. Teachers were informed about the upcoming reports by research assistants while they were visiting the classroom for observations or child assessments. The teachers completed the reports on individual children of the VUOKKO follow-up children in their classroom (1–8 children/classroom) with respect to children's self-regulation skills. In addition, the teachers completed the reports of all the children in their classroom with consent to take part in the study with respect to children's math attitudes and needs for differentiated instruction. Of the participating 38 teachers, 36 teachers (95%) returned the reports. Individual reports were completed for 105 VUOKKO follow-up children and for 406 classmates.

TR – School Phase: Grade 3 (T7)

Teachers' reports on individual children, with an introduction letter attached, were sent out to each classroom at the beginning of May 2023. Teachers were informed about the upcoming reports by research assistants while they were visiting the classroom for observations or child assessments. The teachers completed the reports on individual children of the VUOKKO follow-up children in their classroom (1–15 children/classroom) with respect to children's self-regulation skills. In addition, teachers completed the reports of all the children in their classroom with consent to take part in the study with respect to children's math attitudes and needs for differentiated instruction. Of the participating 49 teachers, 41 teachers (84%) returned the reports. Individual reports were completed for 116 VUOKKO follow-up children and for 456 classmates.

5. Instruments in Data Collection / Measures

Section 5 describes the measures used in the VUOKKO study. The aim is to describe each of the measures in a way that one can get a clear impression of what the measure is about. References are provided to the measures when those are available. Some of the measures were developed in the project but of them the questionnaire items in particular may include similar items to scales used in other research. When possible, the references are made to validation studies or manuals, but sometimes references are given to empirical papers that have used the measures before or other larger scale projects that have used the measures before us. More information is available from the research group, also at item level except when the scales or tasks cannot be distributed due to copyright or other related issues.

5.1. CA – Children's Assessments

Children's assessments were conducted in each timepoint: T1 (2.4 years), T2 (2.9 years), T3 (5.4 years), T4 (5.9 years), T5 (Grade 1), T6 (Grade 2) and T7 (Grade 3). Figures 3–6 provide

an overview of the measures grouped in four domains: 1. Child's mathematical skills prior to school entry, 2. Child's mathematical skills at school age, 3. Child's language and literacy skills, and 4. Child's domain-general cognitive skills. The measures are described in detail below the figures.

FIGURE 3 An overview of the **child's mathematical skills prior to school entry**.

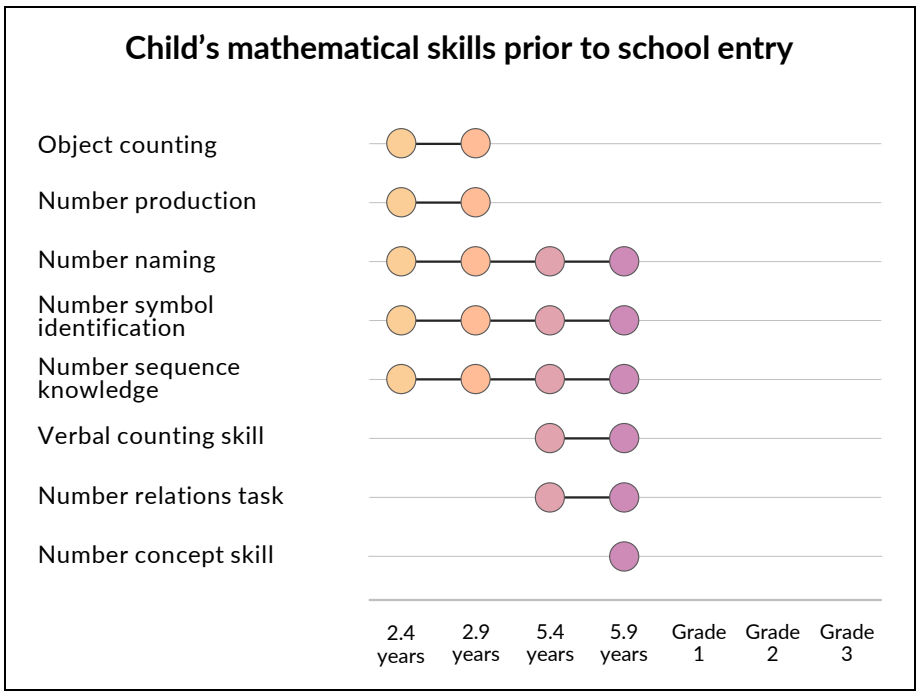
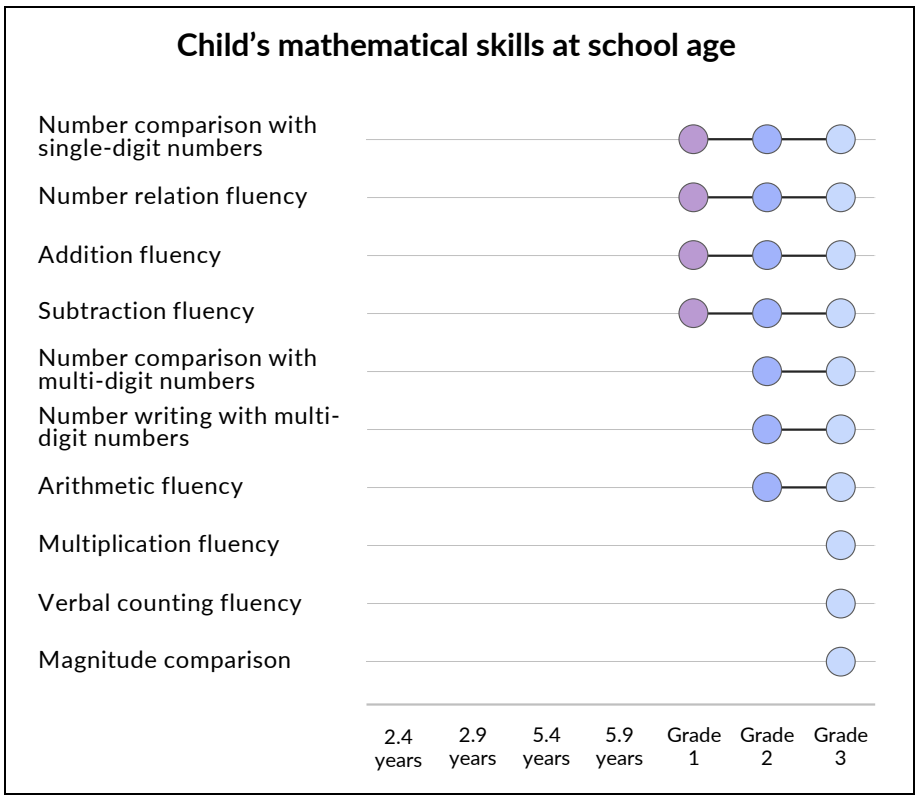


FIGURE 4 An overview of the **child's mathematical skills at school age**.



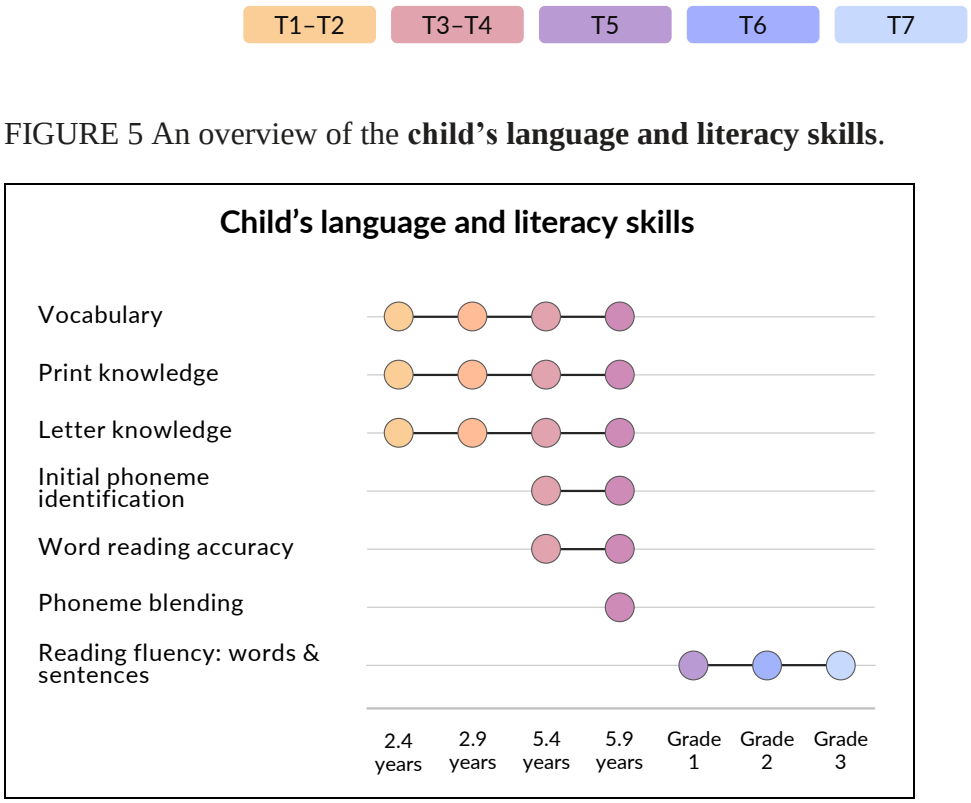


FIGURE 5 An overview of the **child's language and literacy skills**.

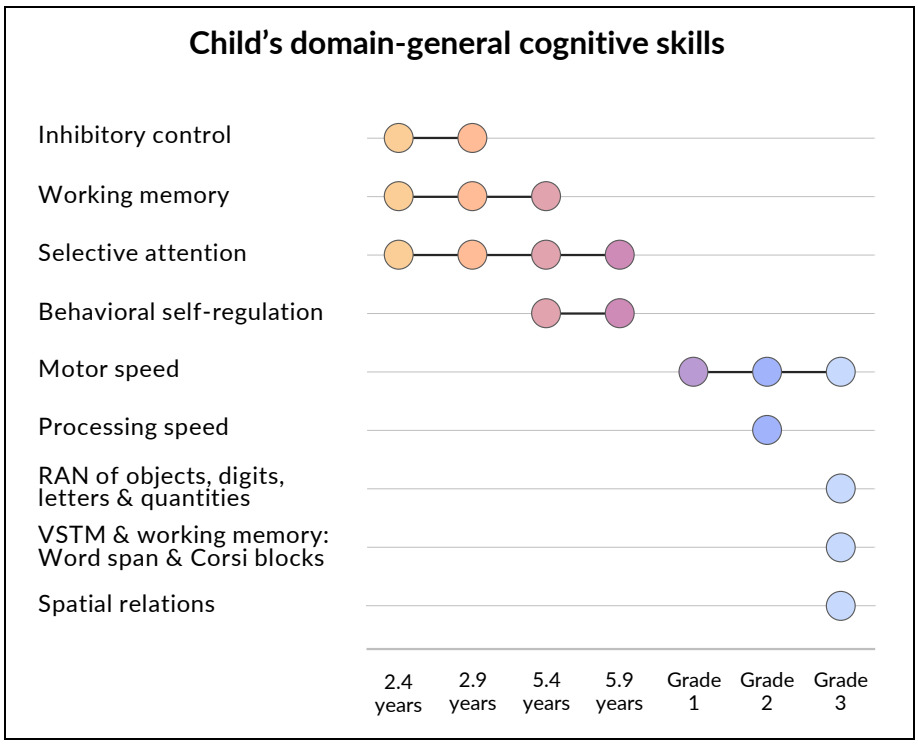


FIGURE 6 An overview of the **child's domain-general cognitive skills**.

5.1.1. CA – Toddler Phase (T1–T2)

Skills were assessed in both fall and spring (T1 and T2, respectively).

NUMERACY SKILLS

Object Counting

The child's skill to enumerate the number of the objects by counting was assessed with an object-counting task modified from Hannula and Lehtinen (2005). The child was asked to count the number of buttons presented. The task included nine test items. The score for the object counting task was the number of correctly counted items (maximum 9 points).

The task began with placing four wooden buttons on the table in front of the child, with a piece of paper blocking the visual field. The paper was removed, and the child was asked to count how many buttons there were. If the answer was correct, the child was credited with one point for this item (4 buttons) and the three preceding items (considered to be able to count 1, 2, and 3 buttons). The test was continued by presenting first 5, then 6, 8, 10, and 12 buttons to be counted. If the child counted four buttons incorrectly, the next trial was two, and if the child further failed on two, the next trial was one button. If the child counted correctly two buttons, the next trial included three buttons. Each item included two trials (the second trial was presented only if the child made an error in the first trial).

Reference:

Hannula, M. M., & Lehtinen, E. (2005). Spontaneous focusing on numerosity and mathematical skills of young children. *Learning and Instruction*, *15*(3), 237–256. https://doi.org/10.1016/j.learninstruc.2005.04.005

Number Production

The "Give me X" task (Wynn, 1990, 1992) was used to tap child's number concept skill. The child's task was to pick up the requested number of figures from a box. The test included eight items. Two alternative scores were composed for the number production task: 1. The number of correct items (maximum 8 points) and 2. the highest number of correctly picked figures (maximum 19 points).

In this task, the child was asked to pick up an amount of plastic figures from a box with a lid on (e.g., "Give me four strawberries"). Before removing the lid and giving the child a turn, the child was asked to confirm how many strawberries they were supposed to give. The task was continued if the child said the correct number word or raised the corresponding number of fingers. In case the child did not respond correctly, the question was repeated. If the child could not respond correctly after repeating the question, the examiner said: "You are supposed to give me four strawberries" and the lid was opened. The test included eight items with increasing difficulty (number of items to be picked were 2, 3, 4, 5, 7, 9, 13, 19), and each item included two trials. If the child failed in both trials for the specific item, the task was terminated.

References:

Wynn, K. (1990). Children's understanding of counting. *Cognition*, *36*, 155–193. https://doi.org/10.1016/0010-0277(90)90003-3

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

Number Naming

Number naming skill was assessed with a task presented by Wright et al. (2006). The child's task was to name numbers, one by one, written in separate cards. The task included 12 items. The number of correct responses was used as the score representing the child's number naming skills (maximum 12 points).

A deck of 12 cards with numbers on them was spread on the table before the child. The cards included the following numbers: 1, 2, 4, 5, 8, 9, 10, 13, 17, 26, 31, 48. One number at a time was pointed at, and the child was asked to say the name of the number. The task proceeded in segments of three numbers with increasing difficulty. If the child knew at least two out of three numbers within a set, three more items were introduced. If not, the task was terminated.

Reference:

Wright, R. J., Martland, J., & Stafford, A. K. (2006). *Early numeracy: Assessment for teaching and intervention.* Sage.

Number Symbol Identification

The child's number symbol identification skill was assessed with a task presented by Wright et al. (2006). The child's task was to point out a requested number from a paper including 10 numbers. The test included 12 items. The number of correct responses was used as the score representing the child's skill in number symbol identification (maximum 12 points).

The 12 test items were grouped into four sets, each including three numbers. The first two number sets included numbers from 1 to 10 on an A4-sized paper. The third set included numbers from 11 to 20, and the fourth set numbers from 21 to 50. The requested numbers were asked in mixed order. If the child knew at least two number symbols within a set, the next set of three numbers was introduced. If the child did not recognize at least two numbers within a set, the task was terminated.

Reference:

Wright, R. J., Martland, J., & Stafford, A. K. (2006). *Early numeracy: Assessment for teaching and intervention*. Sage.

Number Sequence Knowledge

The child's skill in producing number sequences was measured with a verbal counting task (Hannula & Lehtinen, 2005). The child was asked "How high can you count?" The number of correctly produced numbers without mistakes was considered as the score for the child's number sequence skills (maximum 50 points).

The child was encouraged to count as high as they could starting from one. If the child was reluctant or unable to start, the researcher modelled numbers from one to twelve and gave a

turn to the child. The child had two trials, and the longer number sequence without mistakes was scored. The task was terminated if the child reached the maximum score of 50.

Reference:

Hannula, M. M., & Lehtinen, E. (2005). Spontaneous focusing on numerosity and mathematical skills of young children. *Learning and Instruction, 15*(3), 237–256. https://doi.org/10.1016/j.learninstruc.2005.04.005

Language and Literacy Skills

Vocabulary

The child's receptive vocabulary was assessed with the Peabody Picture Vocabulary Test-R (PPVT-Short: Dunn & Dunn, 1981). The child's task was to identify the requested picture out of four alternatives. The test included two practice items and 30 test items. The number of correct responses was used as the score for the child's performance in the test (maximum 30 points).

In this task, the child was shown a set of 30 A4-sized papers with four pictures on each. For each sheet of paper, a tester said a word and the child was asked to point out which picture fitted with the target word. For example, "Point to the picture that has a sock in it." Age-specific items were selected based on the data from a prior longitudinal study, the JLD (Lohvansuu et al., 2021).

References:

Dunn, L., & Dunn, L. (1981). *The Peabody picture vocabulary test-revised*. American Guidance Service.

Lohvansuu, K., Torppa, M., Ahonen, T., Eklund, K., Hämäläinen, J. A., Leppänen, P. H. T., & Lyytinen, H. (2021). Unveiling the mysteries of dyslexia—Lessons learned from the prospective Jyväskylä longitudinal study of dyslexia. *Brain Science, 11, 427.* https://doi.org/10.3390/brainsci11040427

Print Knowledge

The child's print knowledge was assessed with the print awareness subtest of the Test of Preschool Early Literacy (TOPEL; Lonigan et al., 2007). The child's task was to point out the picture with letters on it out of four alternatives. The test included 12 test items. The number of correct responses was used as the score for the print knowledge task (maximum 12 points).

The child was shown a set of 12 A4-sized papers with four pictures on each. Pictures could represent, for example, four different book covers, one with a title written with words, one with a price tag (number symbols), and two with a picture on the cover.

Reference:

Lonigan, C. J., Wagner, R. K., Torgesen, J. K., & Rashotte, C. A. (2007). *TOPEL: Test of preschool early literacy*. ProEd.

Letter Knowledge

Letter knowledge was assessed with the VIIVI test (Lohvansuu et al., 2021; Torppa et al., 2006). The child was asked to name letters written in capitals. The test included 29 letters. The number of correctly named letters was used as the score for the child's letter knowledge (maximum 29 points).

The child was presented one letter at a time on a A4-sized paper. The child was presented with 29 letters organized in four sets (6 + 6 + 4 + 13 letters). The order of the letters is based on the order of their frequency of occurrence in Finnish words and also reflects the ease of pronunciation. The test was terminated if the child did not know any of the letters within a set. The testing always began by presenting the child with the letter expected to be most familiar to the child: the first letter of the child's own first name. The child received one point for each correct response (use of a phoneme or a letter name were both accepted as correct responses).

References:

Lohvansuu, K., Torppa, M., Ahonen, T., Eklund, K., Hämäläinen, J. A., Leppänen, P. H. T., & Lyytinen, H. (2021). Unveiling the mysteries of dyslexia—Lessons learned from the prospective Jyväskylä longitudinal study of dyslexia. *Brain Science*, *11*, 427. https://doi.org/10.3390/brainsci11040427

Torppa, M., Poikkeus, A.-M., Laakso, M. L., Eklund, K., & Lyytinen, H. (2006). Predicting delayed letter knowledge development and its relation to grade 1 reading achievement among children with and without familial risk for dyslexia. *Developmental Psychology, 42*(6), 1128–1142. https://doi.org/10.1037/0012-1649.42.6.1128

Domain-General Cognitive Skills

Inhibitory Control

The child's inhibitory control skill was assessed with the toy wrap task from the Pre-school Self-Regulation Assessment (PSRA; Smith-Donald et al., 2007). The child's task was to avoid peeking and touching a secret and attractive gift. The latency to peek and the latency to touch were used as proxies of the child's inhibitory control skill (maximum 60 seconds)

The researcher started the task by telling the child that they have something nice to play with, but it is a surprise that still needs to be wrapped; the researcher asked the child to look away and not to peek while the gift was being noisily wrapped behind the child's back. The child's latency to peek, that is, the time the child waited until they first peek while the gift is being wrapped, was recorded (maximum 60 seconds). Next, the packed gift was left on the table while the packing material was collected away. The child's latency to touch, that is, the time the child waited until they first touched the gift even though they were told not to touch it, was recorded (maximum 60 seconds).

Reference:

Smith-Donald, R., Raver, C. C., Hayes, T., & Richardson, B. (2007). Preliminary construct and concurrent validity of the pre-school self-regulation assessment (PSRA) for field-based research. *Early Childhood Research Quarterly, 22(2),* 173–187. https://doi.org/10.1016/j.ecresq.2007.01.002

Working Memory

To measure the child's working memory, the hidden toys task (Mulder, et al., 2014) was used. The child's task was to keep in mind which of the six boxes included a toy and pick up a toy from a box when requested. The test included two practice trials with two boxes, which was followed by six test items. The number of toys found was used as a measure representing the child's working memory (maximum 6 points).

For this task, the researcher hid six small toys inside six identical boxes displayed in two rows on a table in front of a child. The child was then asked to find a toy, one by one. Each time the child found a toy, the toy was retrieved, and the box was closed and left empty. Thus, the number of boxes including a toy decreased after each trial. After each trial, the child was distracted for 6 seconds by blocking their visual field before the next trial. To accomplish the task, the child must realize which boxes have already been emptied and which boxes still contain a toy, in addition to holding this information in memory over the delay of distraction.

Reference:

Mulder, H., Hoofs, H., Verhagen, J., van der Veen, I., & Leseman, P. P. (2014). Psychometric properties and convergent and predictive validity of an executive function test battery for two-year-olds. *Frontiers in Psychology, 5*, 733. https://doi.org/10.3389/fpsyg.2014.00733

Selective Attention

The child's selective attention was assessed with the visual attention subtest of the Developmental Neuropsychological Assessment (NEPSY; Korkman et al., 1998). The child's task was to mark the target pictures from a linear array of black and white pictures as quickly and accurately as possible. The age-specific standard score based on the NEPSY manual was used as the score representing the child's selective attention.

The child was asked to scan an array of black and white pictures (altogether 50 pictures on an A3 sheet of paper) and mark the target pictures (altogether 20 pictures) with a stamping pen as quickly and accurately as possible. The task included two trials, and the target pictures were bunnies in the first trial and cats in the second trial. In the bunnies task the pictures were placed diligently on a linear grid, clearly formed by rows and columns, whereas in the cats task, the pictures were scattered more unevenly across the paper. A trial was terminated when the child said they had completed the task or when the time limit of 3 minutes was reached. The number of correct marks, incorrect marks, and time elapsed was recorded separately for both trials. To calculate the standard scores, the sum of the incorrect responses for both trials was subtracted from the sum of correct responses for both trials, and this score was divided by the total time used in both trials. The standard score was then determined with the normative, age-specified tables provided in the NEPSY manual.

Reference:

Korkman, M., Kirk, U., & Kemp, S. L. (1998). *NEPSY. A developmental neuropsychological assessment.* Psychological Corporation.

5.1.2. CA – Preschool Phase (T3–T4)

Skills were assessed both in fall and spring (T3 and T4, respectively), if not told otherwise.

NUMERACY SKILLS

Number Naming

Number naming skill was assessed with a task presented by Wright et al. (2006). The child's task was to name numbers, one by one, written in separate cards. The task included 12 items. The number of correct responses was used as the score representing the child's number naming skills (maximum 12 points).

A deck of 12 cards with numbers on them was spread on the table before the child. The cards included the following numbers: 1, 2, 4, 5, 8, 9, 10, 13, 17, 26, 31, 48. One number at a time was pointed at, and the child was asked to say the name of the number. The task proceeded in segments of three numbers with increasing difficulty. If the child knew at least two out of three numbers within a set, three more items were introduced. If not, the task was terminated.

Reference:

Wright, R. J., Martland, J., & Stafford, A. K. (2006). *Early numeracy: Assessment for teaching and intervention.* Sage.

Number Symbol Identification

The child's number symbol identification skill was assessed with a task presented by Wright et al. (2006). The child's task was to point out a requested number from a paper including 10 numbers. The test included 12 items. The number of correct responses was used as the score representing the child's skill in number symbol identification (maximum 12 points).

The 12 test items were grouped into four sets, each including three numbers. The first two number sets included numbers from 1 to 10 on an A4-sized paper. The third set included numbers from 11 to 20, and the fourth set numbers from 21 to 50. The requested numbers were asked in mixed order. If the child knew at least two number symbols within a set, the next set of three numbers was introduced. If the child did not recognize at least two numbers within a set, the task was terminated.

Reference:

Wright, R. J., Martland, J., & Stafford, A. K. (2006). *Early numeracy: Assessment for teaching and intervention.* Sage.

Number Sequence Knowledge

The child's skill in producing number sequences was measured with a verbal counting task (Hannula & Lehtinen, 2005). The child was asked "How high can you count?" The number of correctly produced numbers without mistakes was considered as the score for the child's number sequence skills (maximum 50 points).

The child was encouraged to count as high as they could starting from one. If the child was reluctant or unable to start, the researcher modelled numbers from one to twelve and gave a turn to the child. The child had two trials, and the longer number sequence without mistakes was scored. The task was terminated if the child reached the maximum score of 50.

Reference:

Hannula, M. M., & Lehtinen, E. (2005). Spontaneous focusing on numerosity and mathematical skills of young children. *Learning and Instruction, 15*(3), 237–256. https://doi.org/10.1016/j.learninstruc.2005.04.005

Verbal Counting Skill

Child's skill to count forward and backward from a given number was measured with a modified version of the counting subtask retrieved from the *LukiMat – Oppimisen Arviointi* screening battery (Koponen et al., 2011). The child was asked to count five steps forward or backward from a given number word. The task included three counting forward items and three counting backward items. The number of correctly counted number sequences was used as the score (maximum 6 points).

The child was first asked to orally count forward three number sequences, in steps of 1, first, from 4 to 8, second from 9 to 13, and finally from 17 to 21. Next, the child was asked to orally count backward number sequences, in steps of 1, first, from 5 to 1, second, from 12 to 8, and third, from 21 to17. Each correctly counted number sequence received one point.

Reference:

Koponen, T., Salminen, J., Aunio, P., & Polet, J. (2011). *LukiMat – Oppimisen arviointi: Matematiikan tuen tarpeen tunnistamisen välineet esikouluun. Käyttäjän opas* [LukiMat – Evaluation of learning: Tools for identifying the need of support in math on pre-primary education. User manual]. Niilo Mäki Institute. http://www.lukimat.fi/lukimat-oppimisen-arviointi/materiaalit/tuen-tarpeen-tunnistaminen/esiopetus/matematiikka/kayttajan-opas

Number Relations Task

The ability to compare magnitude and state the exact numerical difference between two numbers was assessed with a number relations task (Koponen, 2018). The number of the candies in two boxes were expressed with number words (Subtest A) or written number symbols (Subtest B). The child was asked to (a) show/say which of the two boxes included more candies, and then to say (b) how many more. The task included one practice item and six test items for each subtest. For both subtests two scores were composed: the number of correctly selected boxes and number of correctly stated numerical differences (maximum 6 points for each), and thus, four scores altogether.

Two identical boxes were placed on the table in front of the child. They were told that the boxes contained candies. In the practice item boxes were open and the child saw one candy and two candies. The child was asked, first, which of the two boxes had more candies in it, and second, how many more are two candies compared to one candy. In the test items, the number of candies in each box was told aloud (subtest A) or presented with a number symbol (subtest B) and then the child was asked: "Which box has more candies in it?" and "How many more?"

Reference:

Koponen, T. (2018). *Number relations task.* Unpublished task

Number Concept Skill

Child's number concept skill was assessed using a task from a national school readiness test (Elomäki et al., 1999), but only in the spring (T4). The child's task was to follow verbal instructions related to different math concepts. The task included nine test items. The number of correct responses was used as the score representing the child's math concept knowledge (maximum 9 points).

The test consisted of nine subtasks completed in paper and pen format. The child was asked to (1) draw as many balls as was shown in the picture; (2) draw one ball more than in the picture; (3) draw one ball less than in the picture; (4) draw as many balls as the number symbol indicates; (5) circle the number which tells how many balls there are in the box (6) circle the number which tells how many balls there are in the box; (7) mark off the first ball; (8) mark off the fourth ball; and (9) mark off the seventh ball. The child got one point for each correct answer.

Reference:

Elomäki, T., Huolila, R., Poskiparta, E., & Saranpää, P. (1999). *Kouluvalmiuden arviointi ryhmässä. Ryhmätutkimuksen käsikirja ja seurantatutkimus* [Group assessment of school readiness. Handbook of group assessment and a follow-up study] (Publications of the Social and Health Center, City of Turku, Finland: A, 2). Social and Health Center.

Language and Literacy Skills

Vocabulary

The child's receptive vocabulary was assessed with the Peabody Picture Vocabulary Test-R (PPVT-Short: Dunn & Dunn, 1981). The child's task was to identify the requested picture out of four alternatives. The test included two practice items and 30 test items. The number of correct responses was used as the score for the child's performance in the test (maximum 30 points).

In this task, the child was shown a set of 30 A4-sized papers with four pictures on each. For each sheet of paper, a tester said a word and the child was asked to point out which picture fitted with the target word. For example, "Point to the picture that has a sock in it." Age-specific items were selected based on the data from a prior longitudinal study, the First Steps (Lerkkanen et al., 2006).

References:

Dunn, L., & Dunn, L. (1981). *The Peabody picture vocabulary test-revised.* American Guidance Service.

Lerkkanen, M.-K., Niemi, P., Poikkeus, A.-M., Poskiparta, E., Siekkinen, M., & Nurmi, J.-E. (2006–2016). *The First Steps Study* [Alkuportaat]. University of Jyväskylä.

Print Knowledge

The child's print knowledge was assessed with the print awareness subtest of the Test of Preschool Early Literacy (TOPEL; Lonigan et al., 2007). The child's task was to point out the picture with letters on it out of four alternatives. The test included 12 test items. The number of correct responses was used as the score for the print knowledge task (maximum 12 points).

The child was shown a set of 12 A4-sized papers with four pictures on each. Pictures could represent, for example, four different book covers, one with a title written with words, one with a price tag (number symbols), and two with a picture on the cover.

Reference:

Lonigan, C. J., Wagner, R. K., Torgesen, J. K., & Rashotte, C. A. (2007). *TOPEL: Test of preschool early literacy.* ProEd.

Letter Knowledge

Letter knowledge was assessed with the ARMI test (Lerkkanen et al., 2006). The child was asked to name all 29 uppercase letters used in the Finnish alphabet. The number of correctly named letters was used as the score for the child's letter knowledge (maximum 29 points).

In the ARMI test, all 29 uppercase letters of the Finnish alphabet were presented to the child in three rows, one row at a time, printed on an A4 sheet. The child was asked to name aloud all letters in every row, from left to right.

Reference:

Lerkkanen, M.-K., Poikkeus, A.-M., & Ketonen, R. (2006). *ARMI—Luku- ja kirjoitustaidon arviointimateriaali 1. luokalle* [ARMI—A tool for assessing reading and writing skills in grade 1]. WSOY.

Initial Phoneme Identification

Initial phoneme identification was assessed with the ARMI test (Lerkkanen et al., 2006). The child's task was to identify the correct picture with the requested first phoneme out of four alternatives. The task included one practice item and 10 test items. The number of correct items was used as the score representing the child's initial phoneme identification (maximum value of 10).

The child was shown 10 sets of four pictures, one at a time. The four pictures within each set were named. The child was instructed as follows: "Here is a picture of *omena*, *sukka*, *reppu* and *lintu* [apple, sock, bag, bird]. Listen carefully: which of the words starts with the sound /o/: *omena*, *sukka*, *reppu*, *lintu*?"

Reference:

Lerkkanen, M.-K., Poikkeus, A.-M., & Ketonen, R. (2006). *ARMI—Luku- ja kirjoitustaidon arviointimateriaali 1. luokalle* [ARMI—A tool for assessing reading and writing skills in grade 1]. WSOY.

Word Reading Accuracy

Word reading skill was assessed using a wordlist of the ARMI test (Lerkkanen et al., 2006). The child's task was to read aloud words presented in uppercase letters, one at a time. The number of correctly read words was used as the score representing the child's word reading skill (maximum value of 10).

The list included 10 words (7 two-syllabic, 2 three-syllabic, and 1 five-syllabic) with 3 to 13 letters. The words were presented to the child one at a time in the order of increasing difficulty and covering the remaining words with an empty sheet of paper. The task was terminated if the child read incorrectly three items in a row.

Reference:

Lerkkanen, M.-K., Poikkeus, A.-M., & Ketonen, R. (2006). ARMI—Luku- ja kirjoitustaidon arviointimateriaali 1. luokalle [ARMI—A tool for assessing reading and writing skills in grade 1]. WSOY.

Phoneme Blending

The phoneme blending skill was assessed with the phoneme blending test (Poskiparta et al., 1994), but only in the spring (T4). The child's task was to identify the requested picture out of four alternatives, the name of which the tester had sounded phoneme by phoneme. The test consisted of one practice item and nine test items. The number of correct responses was used as the score representing the child's phoneme blending skills (maximum value of 9).

The child was shown four pictures of objects side by side in a row. Next, the research assistant sounded out the phonemes of a word corresponding to one of the four pictures (e.g., /y/ /ö/). The child was asked to point at the picture to which the word just sounded was related (e.g., to a picture of a moon in the sky). The nine target words used included two-syllabic (7 items) and three-syllabic (2 items) words with 3 to 7 phonemes.

Reference:

Poskiparta, E., Niemi, P., & Lepola, J. (1994). *Diagnostiset testit 1. Lukeminen ja kirjoittaminen* [Diagnostic tests 1. Reading and writing]. University of Turku, Centre for Research on Learning.

Domain-General Cognitive Skills

Working Memory

To measure the child's working memory, the hidden toys task (Mulder et al., 2014) was used, but only in the fall (T3). The child's task was to keep in mind which of the six boxes included a toy and pick up a toy from a box when requested. The test included two practice trials with two boxes, which was followed by six test items. The number of toys found was used as a measure representing the child's working memory (maximum 6 points).

For this task, the researcher hid six small toys inside six identical boxes displayed in two rows on a table in front of a child. The child was then asked to find a toy, one by one. Each time the child found a toy, the toy was retrieved, and the box was closed and left empty. Thus, the

number of boxes including a toy decreased after each trial. After each trial, the child was distracted for 6 seconds by blocking their visual field before the next trial. To accomplish the task, the child must realize which boxes have already been emptied and which boxes still contain a toy, in addition to holding this information in memory over the delay of distraction.

Reference:

Mulder, H., Hoofs, H., Verhagen, J., van der Veen, I., & Leseman, P. P. (2014). Psychometric properties and convergent and predictive validity of an executive function test battery for two-year-olds. *Frontiers in Psychology, 5*, 733. https://doi.org/10.3389/fpsyg.2014.00733

Selective Attention

The child's selective attention was assessed with the visual attention subtest of the Developmental Neuropsychological Assessment (NEPSY; Korkman et al., 1998). The child's task was to mark the target pictures from a linear array of black and white pictures as quickly and accurately as possible. The age-specific standard score based on the NEPSY manual was used as the score representing the child's selective attention.

The child was asked to scan a linear array of black and white pictures (altogether 50 pictures on an A3 sheet of paper) and mark the target pictures (altogether 20 pictures) with a stamping pen as quickly and accurately as possible. The task included two trials, and the target pictures were bunnies in the first trial and cats in the second trial. In the bunnies task the pictures were placed diligently on a linear grid, clearly formed by rows and columns, whereas in the cats task, the pictures were scattered more unevenly across the paper. A trial was terminated, when the child said they had completed the task or when the time limit of 3 minutes was reached. The number of correct marks, incorrect marks, and time elapsed was recorded separately for both trials. To calculate the standard scores, the sum of the incorrect responses for both trials was subtracted from the sum of correct responses for both trials, and this score was divided by the total time used in both trials. The standard score was then determined with the normative, age-specified tables provided in the NEPSY manual.

Reference:

Korkman, M., Kirk, U., & Kemp, S. L. (1998). *NEPSY. A developmental neuropsychological assessment.* Psychological Corporation.

Behavioral Self-Regulation

Behavioral self-regulation was measured with the Head-Toes-Knees-Shoulders task (HTKS [*Form-A*], McClelland et al., 2014). The child's task was to inhibit orally given requests and act according to given rules. The task includes 30 test items. Each item was scored with 0 to 2 points: 0 = incorrect, 1 = self-correct, and 2 = correct. Self-correction was defined as any motion to the incorrect response, but self-correcting and ending up with the correct action. The sum of the scores to all 30 items was used as the measure representing the child's ability to behavioral self-regulation (maximum 60 points).

The HTKS requires cognitive flexibility, working memory, and inhibitory control. The test comprises 30 items spread across three sections that become gradually more complex: Section 1 introduces opposites, section 2 two sets of opposites, and section 3 adds a final rule switch.

Each of the three sections had 10 items. After sections 1 and 2 the test was terminated in case the child scored 0 to 3 points in the corresponding section.

In section 1, after habituating to two oral commands (i.e., “touch your head” and “touch your toes”), the child was asked to respond in an unnatural way to two types of paired behavioral commands (e.g., touch their head when told to touch their toes).

In section 2, the child was habituated to two more oral commands (i.e., “touch your knees” and “touch your shoulders”). After that the child was asked to respond in an unnatural way to four types of paired behavioral commands, following the rules habituated in sections 1 and 2 (e.g., touch their head when told to touch their toes; touching their shoulder when told to touch their knees).

In section 3, the child was habituated to four types of paired behavioral commands that did not follow the same rules as the previous two sections (i.e., pairing head with knees and shoulders with toes).

Reference:

McClelland, M. M., Cameron, C. E., Duncan, R., Bowles, R. P., Acock, A. C., Miao, A., & Pratt, M. E. (2014). Predictors of early growth in academic achievement: The head-toes-knees-shoulders task. *Frontiers in psychology, 5*, 81720. https://doi.org/10.3389/fpsyg.2014.00599

5.1.3. CA – School Phase: Grade 1 (T5)

MATHEMATICAL SKILLS

Number Comparison with Single-Digit Numbers

Number comparison skill was assessed using the paper and pencil task where the task was to identify the larger of the two single-digit numbers (from one to nine) (Brankaer et al., 2017). A maximum of 60 items could be attempted within a 45 second time limit. One point was given from each correct answer. The sum of correct answers within the time limit was used as the score for the number comparison task (maximum 60). The numbers of attempted, incorrect and missing answers (items which a child had skipped) were coded as separate variables in the data.

The task included a front page and one test sheet. The front page had two examples and two practice items. The test sheet had four columns of which each had 15 items (a box with two numbers).

The children were asked to cross out the number that was greater than the adjacent number in the same box. First, the children were asked to complete the practice items. The practice items were quickly checked together with the class prior to giving them permission to continue to the actual task. In the actual task the children were asked to complete as many items as fast and accurately as possible in the given time. The children were told to proceed one column at a time from top to bottom and after completing a column to proceed to the following one. They were also instructed to correct a mistake if they made one by smudging the incorrect line and then selecting the correct number.

Reference:

Brankaer, C., Ghesquière, P., & De Smedt, B. (2017). Symbolic magnitude processing in elementary school children: A group administered paper-and-pencil measure (SYMP Test). *Behavior Research Methods*, *49*, 1361–1373. https://doi.org/10.3758/s13428-016-0792-3

Number Relation Fluency

The ability to detect exact numerical differences between two numbers was assessed by the number relation fluency task (Koponen, 2021b). Children were asked to mark how much larger one of the numbers was in comparison with the other number. There was a 60-seconds time limit. The number range varied between one and ten and the difference in magnitude varied from one to three. One point was given from each correct answer and the sum of correct answers within the time limit was used as the score for the task (maximum 40 points). The numbers of attempted, incorrect and missing answers (items which a child had skipped) were coded as separate variables in the data.

The task included a front page and one test sheet. The front page had two examples and two practice items. The test sheet had four columns with 10 items on each (two numbers presented next to each other and an empty answer box below them).

The children were asked to mark how much larger one of the numbers was in comparison with the other number in the item. First, the children were asked to complete the practice items. The practice items were quickly checked together with the class prior to giving them permission to continue to the actual task. In the actual task the children were asked to complete as many items as fast and accurately as possible in the given time. The children were told to proceed one column at a time from top to bottom and after completing a column to proceed to the following one. They were also instructed to correct a mistake if they made one by smudging the incorrect number and drawing the correct one next to it.

Reference:

Koponen, T. (2021b). *Number relation fluency test*. Unpublished task.

Addition Fluency

In the addition fluency task (Koponen & Mononen, 2010a) children were asked to calculate as many single-digit additions (e.g., 1 + 3 = __; 9 + 5 = __) out of 120 items as possible within a 2-minute time limit. One point was given for each correct answer and the task score is the sum of correct answers within the time limit (maximum 120). The numbers of attempted, incorrect and missing answers (items which a child had skipped) were coded as separate variables in the data.

The task included a front page and two test sheets. The front page had two examples and two practice items. The two test sheets had three columns with 20 items on each.

The children were asked to calculate the addition and write the answer next to each problem. First, the children were asked to complete the practice items. The practice items were quickly checked together with the class prior to giving them permission to continue to the actual task. In the actual task the children were asked to calculate as fast and accurately as possible in the given time. The children were told to proceed one column at a time from top to bottom and

after completing a column to continue to the following one. In addition, the children were asked to turn the page immediately if they completed the first page. They were also instructed to correct a mistake if they made one by smudging the incorrect number and drawing the correct one next to it.

Reference:

Koponen, T., & Mononen, R. (2010a). *The 2-minute addition fluency test*. Unpublished task.

Subtraction Fluency

In the subtraction fluency task (Koponen & Mononen, 2010b) children were asked to calculate as many subtractions (e.g., 5 – 2 = __; 12 – 8 = __) out of 120 items as possible within a 2-minute time limit. Altogether 120 items with a minuend below 20 and a single-digit subtrahend were presented. One point was given from each correct answer and the task score is the sum of correct answers within the time limit (maximum 120). The numbers of attempted, incorrect and missing answers (items which a child had skipped) were coded as separate variables in the data.

The task included a front page and two test sheets. The front page had two examples and two practice items. The two test sheets had three columns with 20 items on each.

The children were asked to calculate the subtraction and write the answer right next to each problem. First, the children were asked to complete the practice items. The practice items were quickly checked together with the class prior to giving them permission to continue to the actual task. In the actual task the children were asked to calculate as fast and accurately as possible in the given time. The children were told to proceed one column at a time from top to bottom and after completing a column to continue to the following one. In addition, the children were asked to turn the page immediately if they completed the first page. They were also instructed to correct a mistake if they made one by smudging the incorrect number and drawing the correct one next to it.

Reference:

Koponen, T., & Mononen, R. (2010b). *The 2-minute subtraction fluency test*. Unpublished task.

Reading Skills

Reading Fluency: Word Reading

A subtest (ALLU TL2: version A) of the nationally standardized reading test battery (Lindeman, 1998) was used to assess word-level reading fluency and accuracy. Each item contained a picture with four words next to it. The task was to read the four phonologically similar words as fast as possible and draw a line connecting each picture to the word that matched it within the 2-minute time limit. The number of correct items within the time limit was used as the score representing the children's word reading fluency (maximum 80). The numbers of attempted, incorrect and missing answers (items which a child had skipped) were coded as separate variables in the data.

The task material included a front page and eight test sheets. The front page had four practice items, and each of the test sheets had five rows of which each had two items.

The children were asked to read words as fast as possible and to draw a line connecting each picture to the correct word. The children were encouraged to guess the correct answer in case they could not yet read the words. First, the children were asked to complete the practice items. The practice items were quickly checked together with the class prior to giving them permission to continue to the actual task. In the actual task the children were asked to complete as many items as fast and accurately as possible in the given time. The children were told to proceed one row at a time from left to right, and to turn the next page immediately if they completed the previous page. They were instructed to move on to the following item if they found some of them too challenging. They were also instructed to correct a mistake if they made one by smudging the incorrect line and drawing a new line from the picture to the correct word.

Reference:

Lindeman, J. (1998). *ALLU—Ala-asteen lukutesti* [Reading test for primary school]. University of Turku.

Reading Fluency: Sentence Reading

The TOSREC test of silent reading efficiency and comprehension was used as a measure of sentence-level reading fluency and accuracy (Wagner et al., 2009). In the test children read and evaluated the truthfulness of sentences based on real-world knowledge (e.g., “You can look at the book” and “An elephant is small”). The task was to verify the truthfulness of as many sentences as possible out of 60 items within the 3-minute time limit. The number of correct items within the time limit was used as the score representing the children’s sentence reading fluency (maximum 60). The numbers of attempted, incorrect and missing answers (items which a child had skipped) were coded as separate variables in the data.

The task material included a front page and three test sheets. The front page had two examples and five practice items, and each of the test sheets had 20 items that were placed one below another.

The children were asked to read each sentence fast and carefully and think whether the sentence is true or not. They were instructed to circle the word “Yes,” if the sentence was true, and to circle “No,” if the sentence was false. Sentences were constructed in a way that the truthfulness of the sentence would be easy to decide, and therefore, the variation in scores between individuals would mainly reflect their differences in reading fluency. First, the children were asked to complete the practice items. The practice items were quickly checked together with the class prior to giving them permission to continue to the actual task. In the actual task the children were asked to complete as many items as accurately as possible in the given time. The children were told to proceed one sheet at a time from top to bottom and to turn the page immediately if they completed the previous page. They were instructed to move on to the following item if they found some of them too challenging. They were also instructed to correct a mistake if they made one by smudging the incorrect answer and circling the correct word.

Reference:

Wagner, R. K., Torgesen, J. K., Rashotte, C. A., & Pearson, N. A. (2009). *TOSREC: Test of sentence reading efficiency and comprehension.* Pro-Ed.

Domain-General Cognitive Skills

Motor Speed (Copying Numbers)

Motor speed was assessed using a number copying task (Koponen, 2020). The task was to copy a given number (between 1 and 10) as fast and accurately as possible. The task included 90 items and had a 45-second time limit. The number of correct items within the time limit was used as the children's score representing motor speed (maximum 90 points). The numbers of attempted, incorrect and missing answers (items which a child had skipped) were coded as separate variables in the data.

The task Included a front page and two test sheets. The front page had two examples and two practice items, and the test sheets had 45 items arranged in three columns. Each column had 15 items.

The children were asked to copy the given number right next to each number. First, the children were asked to complete the practice items. The practice items were quickly checked together with the class prior to giving them permission to continue to the actual task. In the actual task the children were asked to complete as many items as fast and accurately as possible in the given time. The children were told to proceed one column at a time from top to bottom and after completing a column to continue to the following one. They were also instructed to turn the page immediately if they completed the first page. In addition, the children were instructed to correct a mistake if they made one by smudging the incorrect number and drawing the correct one next to it.

Reference:

Koponen, T. (2020). *Motor speed-copying digits.* Unpublished task.

5.1.4. CA – School Phase: Grade 2 (T6)

Mathematical Skills

Number Comparison with Single-Digit Numbers

Number comparison skill was assessed using the paper and pencil task where the task was to identify the larger of the two single-digit numbers (from one to nine) (Brankaer et al., 2017). A maximum of 60 items could be attempted within a 45 second time limit. One point was given from each correct answer. The sum of correct answers within the time limit was used as the score for the number comparison task (maximum 60). The numbers of attempted, incorrect and missing answers (items which a child had skipped) were coded as separate variables in the data.

The task included a front page and one test sheet. The front page had two examples and two practice items. The test sheet had four columns of which each had 15 items (a box with two numbers).

The children were asked to cross out the number that was greater than the adjacent number in the same box. First, the children were asked to complete the practice items. The practice items were quickly checked together with the class prior to giving them permission to continue to the

actual task. In the actual task the children were asked to complete as many items as fast and accurately as possible in the given time. The children were told to proceed one column at a time from top to bottom and after completing a column to proceed to the following one. They were also instructed to correct a mistake if they made one by smudging the incorrect line and then selecting the correct number.

Reference:

Brankaer, C., Ghesquière, P., & De Smedt, B. (2017). Symbolic magnitude processing in elementary school children: A group administered paper-and-pencil measure (SYMP Test). *Behavior Research Methods, 49*, 1361–1373. https://doi.org/10.3758/s13428-016-0792-3

Number Relation Fluency

The ability to detect exact numerical differences between two numbers was assessed by the number relation fluency task (Koponen, 2021b). Children were asked to mark how much larger one of the numbers was in comparison with the other number. There was a 60-seconds time limit. The number range varied between one and ten and the difference in magnitude varied from one to three. One point was given from each correct answer and the sum of correct answers within the time limit was used as the score for the task (maximum 40 points). The numbers of attempted, incorrect and missing answers (items which a child had skipped) were coded as separate variables in the data.

The task included a front page and one test sheet. The front page had two examples and two practice items. The test sheet had four columns with 10 items on each (two numbers presented next to each other and an empty answer box below them).

The children were asked to mark how much larger one of the numbers was in comparison with the other number in the item. First, the children were asked to complete the practice items. The practice items were quickly checked together with the class prior to giving them permission to continue to the actual task. In the actual task the children were asked to complete as many items as fast and accurately as possible in the given time. The children were told to proceed one column at a time from top to bottom and after completing a column to proceed to the following one. They were also instructed to correct a mistake if they made one by smudging the incorrect number and drawing the correct one next to it.

Reference:

Koponen, T. (2021b). *Number relation fluency test*. Unpublished task.

Addition Fluency

In the addition fluency task (Koponen & Mononen, 2010a) children were asked to calculate as many single-digit additions (e.g., 1 + 3 = __; 9 + 5 = __) out of 120 items as possible within a 2-minute time limit. One point was given for each correct answer and the task score is the sum of correct answers within the time limit (maximum 120). The numbers of attempted, incorrect and missing answers (items which a child had skipped) were coded as separate variables in the data.

The task included a front page and two test sheets. The front page had two examples and two practice items. The two test sheets had three columns with 20 items on each.

The children were asked to calculate the addition and write the answer next to each problem. First, the children were asked to complete the practice items. The practice items were quickly checked together with the class prior to giving them permission to continue to the actual task. In the actual task the children were asked to calculate as fast and accurately as possible in the given time. The children were told to proceed one column at a time from top to bottom and after completing a column to continue to the following one. In addition, the children were asked to turn the page immediately if they completed the first page. They were also instructed to correct a mistake if they made one by smudging the incorrect number and drawing the correct one next to it.

Reference:

Koponen, T., & Mononen, R. (2010a). *The 2-minute addition fluency test*. Unpublished task.

Subtraction Fluency

In the subtraction fluency task (Koponen & Mononen, 2010b) children were asked to calculate as many subtractions (e.g., 5 – 2 = __; 12 – 8 = __) out of 120 items as possible within a 2-minute time limit. Altogether 120 items with a minuend below 20 and a single-digit subtrahend were presented. One point was given from each correct answer and the task score is the sum of correct answers within the time limit (maximum 120). The numbers of attempted, incorrect and missing answers (items which a child had skipped) were coded as separate variables in the data.

The task included a front page and two test sheets. The front page had two examples and two practice items. The two test sheets had three columns with 20 items on each.

The children were asked to calculate the subtraction and write the answer right next to each problem. First, the children were asked to complete the practice items. The practice items were quickly checked together with the class prior to giving them permission to continue to the actual task. In the actual task the children were asked to calculate as fast and accurately as possible in the given time. The children were told to proceed one column at a time from top to bottom and after completing a column to continue to the following one. In addition, the children were asked to turn the page immediately if they completed the first page. They were also instructed to correct a mistake if they made one by smudging the incorrect number and drawing the correct one next to it.

Reference:

Koponen, T., & Mononen, R. (2010b). *The 2-minute subtraction fluency test*. Unpublished task.

Number Comparison with Multi-Digit Numbers

Multi-digit number comparison was assessed using a paper and pencil task (Koponen, 2018b). The children were asked to identify the largest number out of four numbers written on one row. The task included 20 items that became gradually more difficult: the four number series ranged from single-digit numbers (e.g., 6, 8, 3, 6) to five-digits numbers (e.g., 32999, 60066, 55655, 60504). One point was given from each correct answer and the sum of correct answers was used as the score for the task (maximum 20 points). The numbers of attempted, incorrect and missing answers (items which a child had skipped) were coded as separate variables in the data.

The task included one sheet that had one practice item and 20 test items, one below another. Each item had a letter at the beginning of the row to indicate separate items, followed by four numbers, that were side by side on separate boxes of equal sizes.

The children were asked to cross out the largest number out of four adjacent numbers. One example item was checked together with the children prior to giving them permission to continue to the actual task. In the actual task the children were instructed to work as carefully as they could and use the time they needed. They were encouraged to choose the number that they thought might be the largest even in case they were not sure about the correct answer. They were also instructed to correct a mistake if they made one by smudging the incorrect line and then selecting the correct number.

Reference:

Koponen, T. (2018b). *Number comparison: multidigit numbers*. Unpublished task.

Number Writing with Multi-Digit Numbers

Number writing skill was assessed using the task where the children were asked to write numbers that the researcher dictated one at a time (Koponen, 2018a). The task included 16 items and the numbers ranged from two-digits to five-digits. One point was given from each correctly written number and the sum of correct answers was used as the task score (maximum 16 points). The numbers of attempted, incorrect and missing answers (items which a child had skipped) were coded as separate variables in the data.

The task had one sheet that had empty lines for one practice item and 16 test items, one below another. Each item had a different letter at the beginning of the row to make it easier to follow the instruction, that is, to follow on which line the number related to each item was supposed to be written.

The children were instructed to write numbers as dictated by the researcher. Each number was dictated once by default but repeated up to two times if necessary (altogether three times). The children were encouraged to write the number as they thought it might be even though they weren't sure how to write it. The children were instructed to work as carefully as they could and follow the researcher's guidance. They were also instructed to correct a mistake if they made one by smudging the incorrect number and drawing the correct one next to it.

Reference:

Koponen, T. (2018a). *Writing numbers*. Unpublished task.

Arithmetic Fluency

Arithmetic calculation fluency was assessed with the Basic Arithmetic Test (Aunola & Räsänen, 2007). The task was to calculate single-digit and multi-digit additions (e.g., 3 + 2 = __; 7 + __ = 14) and subtractions (e.g., __ – 3 = 10; 20 – 4 – 3 = __) as accurately and fast as possible within a 3-minute time limit. There were 28 items, and one point was given from each correct answer and the sum of correct answers within the time limit was used as the score for the task (maximum 28). The numbers of attempted, incorrect and missing answers (items which a child had skipped) were coded as separate variables in the data.

The task included a front page and two test sheets. The front page had six practice items. The test sheets had seven rows of which each had two items.

The children were asked to do the calculations and write the answer on the empty line. First, the children were asked to complete the practice items. The practice items were quickly checked together with the class prior to giving them permission to continue to the actual task. In the actual task the children were asked to complete as many items as fast and accurately as possible in the given time. They were instructed to move on to the following item if they found some of them too challenging. The children were told to proceed one row at a time from left to right, and to turn the page immediately if they completed the first page. They were also instructed to correct a mistake if they made one by smudging the incorrect number and drawing the correct one next to it.

Reference:

Aunola, K., & Räsänen, P. (2007). *The Basic Arithmetic Test*. Unpublished task.

Reading Skills

Reading Fluency: Word Reading

A subtest (ALLU TL2: version B) of the nationally standardized reading test battery (Lindeman, 1998) was used to assess word-level reading fluency and accuracy. Each item contained a picture with four words next to it. The task was to read the four phonologically similar words as fast as possible and draw a line connecting each picture to the word that matched it within the 2-minute time limit. The number of correct items within the time limit was used as the score representing the children's word reading fluency (maximum 80). The numbers of attempted, incorrect and missing answers (items which a child had skipped) were coded as separate variables in the data.

The task material Ied a front page and eight test sheets. The front page had four practice items, and each of the test sheets had five rows of which each had two items.

The children were asked to read words as fast as possible and to draw a line connecting each picture to the correct word. The children were encouraged to guess the correct answer in case they could not yet read the words. First, the children were asked to complete the practice items. The practice items were quickly checked together with the class prior to giving them permission to continue to the actual task. In the actual task the children were asked to complete as many items as fast and accurately as possible in the given time. The children were told to proceed one row at a time from left to right, and to turn the next page immediately if they completed the previous page. They were instructed to move on to the following item if they found some of them too challenging. They were also instructed to correct a mistake if they made one by smudging the incorrect line and drawing a new line from the picture to the correct word.

Reference:

Lindeman, J. (1998). *ALLU—Ala-asteen lukutesti* [Reading test for primary school]. University of Turku.

Reading Fluency: Sentence Reading

The TOSREC test of silent reading efficiency and comprehension was used as a measure of sentence-level reading fluency and accuracy (Wagner et al., 2009). In the test children read and evaluated the truthfulness of sentences based on real-world knowledge (e.g., "You can look at the book" and "An elephant is small"). The task was to verify the truthfulness of as many sentences as possible out of 60 items within the 3-minute time limit. The number of correct items within the time limit was used as the score representing the children's sentence reading fluency (maximum 60). The numbers of attempted, incorrect and missing answers (items which a child had skipped) were coded as separate variables in the data.

The task material included a front page and three test sheets. The front page had two examples and five practice items, and each of the test sheets had 20 items that were placed one below another.

The children were asked to read each sentence fast and carefully and think whether the sentence is true or not. They were instructed to circle the word "Yes," if the sentence was true, and to circle "No," if the sentence was false. Sentences were constructed in a way that the truthfulness of the sentence would be easy to decide, and therefore, the variation in scores between individuals would mainly reflect their differences in reading fluency. First, the children were asked to complete the practice items. The practice items were quickly checked together with the class prior to giving them permission to continue to the actual task. In the actual task the children were asked to complete as many items as accurately as possible in the given time. The children were told to proceed one sheet at a time from top to bottom and to turn the page immediately if they completed the previous page. They were instructed to move on to the following item if they found some of them too challenging. They were also instructed to correct a mistake if they made one by smudging the incorrect answer and circling the correct word.

Reference:

Wagner, R. K., Torgesen, J. K., Rashotte, C. A., & Pearson, N. A. (2009). *TOSREC: Test of sentence reading efficiency and comprehension.* Pro-Ed.

Domain-General Cognitive Skills

Motor Speed (Copying Numbers)

Motor speed was assessed using a number copying task (Koponen, 2020). The task was to copy a given number (between 1 and 10) as fast and accurately as possible. The task included 90 items and had a 45-second time limit. The number of correct items within the time limit was used as the children's score representing motor speed (maximum 90 points). The numbers of attempted, incorrect and missing answers (items which a child had skipped) were coded as separate variables in the data.

The task Included a front page and two test sheets. The front page had two examples and two practice items, and the test sheets had 45 items arranged in three columns. Each column had 15 items.

The children were asked to copy the given number right next to each number. First, the children were asked to complete the practice items. The practice items were quickly checked together

with the class prior to giving them permission to continue to the actual task. In the actual task the children were asked to complete as many items as fast and accurately as possible in the given time. The children were told to proceed one column at a time from top to bottom and after completing a column to continue to the following one. They were also instructed to turn the page immediately if they completed the first page. In addition, the children were instructed to correct a mistake if they made one by smudging the incorrect number and drawing the correct one next to it.

Reference:

Koponen, T. (2020). *Motor speed-copying digits.* Unpublished task.

Processing Speed

Processing speed was assessed using an object identification task (Kail & Ferrer, 2007). The children's attention was drawn to the first object in the row of 20 objects, and their task was to identify five similar objects out of the remaining 19 objects in the row. A maximum of 30 items could be attempted within a 2-minute time limit. One point was given from each test item in which the children had crossed out all five target objects. The number of correct items was used as the score representing the children's processing speed (30 maximum). The numbers of attempted, incorrect and missing answers (items which a child had skipped) were coded as separate variables in the data.

The task included a front page and two test sheets. The front page had altogether six practice items, and each test sheet had 15 test items one below another. Each item had 20 black-and-white objects, (e.g., triangles with varying details) displayed in one row.

The children's attention was first drawn to the first object on the row and then instructed to identify and cross out five similar objects on the same row. The first two practice items were completed together with the researcher, and the next four practice items the children were asked to complete themselves. The practice items were quickly checked together with the class prior to giving them permission to continue to the actual task. In the actual task, the children were asked to complete as many items as fast and accurately as possible in the given time. They were instructed to move on to the following item if they found some of them too challenging. The children were told to proceed one sheet at a time from top to bottom, and to turn the page immediately if they completed the first page. They were also instructed to correct a mistake if they made one by smudging the incorrect line and crossing out the correct objects.

Reference:

Kail, R. V., & Ferrer, E. (2007). Processing speed in childhood and adolescence: Longitudinal models for examining developmental change. *Child development, 78*(6), 1760–1770. https://doi.org/10.1111/j.1467-8624.2007.01088.x

5.1.5. CA – School Phase: Grade 3, Group Assessment (T7)

MATHEMATICAL SKILLS

Number Comparison with Single-Digit Numbers

Number comparison skill was assessed using the paper and pencil task where the task was to identify the larger of the two single-digit numbers (from one to nine) (Brankaer et al., 2017). A maximum of 60 items could be attempted within a 45 second time limit. One point was given from each correct answer. The sum of correct answers within the time limit was used as the score for the number comparison task (maximum 60). The numbers of attempted, incorrect and missing answers (items which a child had skipped) were coded as separate variables in the data.

The task included a front page and one test sheet. The front page had two examples and two practice items. The test sheet had four columns of which each had 15 items (a box with two numbers).

The children were asked to cross out the number that was greater than the adjacent number in the same box. First, the children were asked to complete the practice items. The practice items were quickly checked together with the class prior to giving them permission to continue to the actual task. In the actual task the children were asked to complete as many items as fast and accurately as possible in the given time. The children were told to proceed one column at a time from top to bottom and after completing a column to proceed to the following one. They were also instructed to correct a mistake if they made one by smudging the incorrect line and then selecting the correct number.

Reference:

Brankaer, C., Ghesquière, P., & De Smedt, B. (2017). Symbolic magnitude processing in elementary school children: A group administered paper-and-pencil measure (SYMP Test). *Behavior Research Methods, 49*, 1361–1373. https://doi.org/10.3758/s13428-016-0792-3

Number Relation Fluency

The ability to detect exact numerical differences between two numbers was assessed by the number relation fluency task (Koponen, 2021b). Children were asked to mark how much larger one of the numbers was in comparison with the other number. There was a 60-seconds time limit. The number range varied between one and ten and the difference in magnitude varied from one to three. One point was given from each correct answer and the sum of correct answers within the time limit was used as the score for the task (maximum 40 points). The numbers of attempted, incorrect and missing answers (items which a child had skipped) were coded as separate variables in the data.

The task included a front page and one test sheet. The front page had two examples and two practice items. The test sheet had four columns with 10 items on each (two numbers presented next to each other and an empty answer box below them).

The children were asked to mark how much larger one of the numbers was in comparison with the other number in the item. First, the children were asked to complete the practice items. The

practice items were quickly checked together with the class prior to giving them permission to continue to the actual task. In the actual task the children were asked to complete as many items as fast and accurately as possible in the given time. The children were told to proceed one column at a time from top to bottom and after completing a column to proceed to the following one. They were also instructed to correct a mistake if they made one by smudging the incorrect number and drawing the correct one next to it.

Reference:

Koponen, T. (2021b). *Number relation fluency test*. Unpublished task.

Addition Fluency

In the addition fluency task (Koponen & Mononen, 2010a) children were asked to calculate as many single-digit additions (e.g., 1 + 3 = __; 9 + 5 = __) out of 120 items as possible within a 2-minute time limit. One point was given for each correct answer and the task score is the sum of correct answers within the time limit (maximum 120). The numbers of attempted, incorrect and missing answers (items which a child had skipped) were coded as separate variables in the data.

The task included a front page and two test sheets. The front page had two examples and two practice items. The two test sheets had three columns with 20 items on each.

The children were asked to calculate the addition and write the answer next to each problem. First, the children were asked to complete the practice items. The practice items were quickly checked together with the class prior to giving them permission to continue to the actual task. In the actual task the children were asked to calculate as fast and accurately as possible in the given time. The children were told to proceed one column at a time from top to bottom and after completing a column to continue to the following one. In addition, the children were asked to turn the page immediately if they completed the first page. They were also instructed to correct a mistake if they made one by smudging the incorrect number and drawing the correct one next to it.

Reference:

Koponen, T., & Mononen, R. (2010a). *The 2-minute addition fluency test*. Unpublished task.

Subtraction Fluency

In the subtraction fluency task (Koponen & Mononen, 2010b) children were asked to calculate as many subtractions (e.g., 5 – 2 = __; 12 – 8 = __) out of 120 items as possible within a 2-minute time limit. Altogether 120 items with a minuend below 20 and a single-digit subtrahend were presented. One point was given from each correct answer and the task score is the sum of correct answers within the time limit (maximum 120). The numbers of attempted, incorrect and missing answers (items which a child had skipped) were coded as separate variables in the data.

The task included a front page and two test sheets. The front page had two examples and two practice items. The two test sheets had three columns with 20 items on each.

The children were asked to calculate the subtraction and write the answer right next to each problem. First, the children were asked to complete the practice items. The practice items were quickly checked together with the class prior to giving them permission to continue to the actual

task. In the actual task the children were asked to calculate as fast and accurately as possible in the given time. The children were told to proceed one column at a time from top to bottom and after completing a column to continue to the following one. In addition, the children were asked to turn the page immediately if they completed the first page. They were also instructed to correct a mistake if they made one by smudging the incorrect number and drawing the correct one next to it.

Reference:

Koponen, T., & Mononen, R. (2010b). *The 2-minute subtraction fluency test*. Unpublished task.

Number Comparison with Multi-Digit Numbers

Multi-digit number comparison was assessed using a paper and pencil task (Koponen, 2018b). The children were asked to identify the largest number out of four numbers written on one row. The task included 20 items that became gradually more difficult: the four number series ranged from single-digit numbers (e.g., 6, 8, 3, 6) to five-digits numbers (e.g., 32999, 60066, 55655, 60504). One point was given from each correct answer and the sum of correct answers was used as the score for the task (maximum 20 points). The numbers of attempted, incorrect and missing answers (items which a child had skipped) were coded as separate variables in the data.

The task included one sheet that had one practice item and 20 test items, one below another. Each item had a letter at the beginning of the row to indicate separate items, followed by four numbers, that were side by side on separate boxes of equal sizes.

The children were asked to cross out the largest number out of four adjacent numbers. One example item was checked together with the children prior to giving them permission to continue to the actual task. In the actual task the children were instructed to work as carefully as they could and use the time they needed. They were encouraged to choose the number that they thought might be the largest even in case they were not sure about the correct answer. They were also instructed to correct a mistake if they made one by smudging the incorrect line and then selecting the correct number.

Reference:

Koponen, T. (2018b). *Number comparison: multidigit numbers*. Unpublished task.

Number Writing with Multi-Digit Numbers

Number writing skill was assessed using the task where the children were asked to write numbers that the researcher dictated one at a time (Koponen, 2018a). The task included 16 items and the numbers ranged from two-digits to five-digits. One point was given from each correctly written number and the sum of correct answers was used as the task score (maximum 16 points). The numbers of attempted, incorrect and missing answers (items which a child had skipped) were coded as separate variables in the data.

The task included one sheet that had empty lines for one practice item and 16 test items, one below another. Each item had a different object at the beginning of the row to make it easier to follow the instruction, that is, to follow on which line the number related to each item was supposed to be written.

The children were instructed to write numbers as dictated by the researcher. Each number was dictated once by default but repeated up to two times if necessary (altogether three times). The children were encouraged to write the number as they thought it might be even though they weren't sure how to write it. The children were instructed to work as carefully as they could and follow the researcher's guidance. They were also instructed to correct a mistake if they made one by smudging the incorrect number and drawing the correct one next to it.

Reference:

Koponen, T. (2018a). *Writing numbers*. Unpublished task.

Arithmetic Fluency

Arithmetic calculation fluency was assessed with the Basic Arithmetic Test (Aunola & Räsänen, 2007). The task was to calculate single-digit and multi-digit additions (e.g., 3 + 2 = __; 7 + __ = 14) and subtractions (e.g., __ – 3 = 10; 20 – 4 – 3 = __) as accurately and fast as possible within a 3-minute time limit. There were 28 items, and one point was given from each correct answer and the sum of correct answers within the time limit was used as the score for the task (maximum 28). The numbers of attempted, incorrect and missing answers (items which a child had skipped) were coded as separate variables in the data.

The task included a front page and two test sheets. The front page had six practice items. The test sheets had seven rows of which each had two items.

The children were asked to do the calculations and write the answer on the empty line. First, the children were asked to complete the practice items. The practice items were quickly checked together with the class prior to giving them permission to continue to the actual task. In the actual task the children were asked to complete as many items as fast and accurately as possible in the given time. They were instructed to move on to the following item if they found some of them too challenging. The children were told to proceed one row at a time from left to right, and to turn the page immediately if they completed the first page. They were also instructed to correct a mistake if they made one by smudging the incorrect number and drawing the correct one next to it.

Reference:

Aunola, K., & Räsänen, P. (2007). *The Basic Arithmetic Test*. Unpublished task.

Multiplication Fluency

In the multiplication fluency task (Koponen & Mononen, 2010c) children were asked to calculate as many single-digit multiplications (e.g., 1 x 3 = __; 8 x 6 = __) out of 120 items as possible within a 2-minute time limit. One point was given from each correct answer and the sum of correct answers within the time limit was used as the score for the task (maximum 120). The numbers of attempted, incorrect and missing answers (items which a child had skipped) were coded as separate variables in the data.

The task included a front page and two test sheets. The front page had two examples and two practice items. The two test sheets had three columns with 20 items on each.

The children were asked to calculate the multiplication and write the answer next to each problem. First, the children were asked to complete the practice items. The practice items were

quickly checked together with the class prior to giving them permission to continue to the actual task. In the actual task the children were asked to calculate as fast and accurately as possible in the given time. The children were instructed to proceed one column at a time from top to bottom and after completing a column to continue to the following one. They were also instructed to turn the page immediately if they completed the first page. In addition, the children were instructed to correct a mistake if they made one by smudging the incorrect number and drawing the correct one next to it.

Reference:

Koponen, T., & Mononen, R. (2010c). *The 2-minute multiplication fluency tes*t. Unpublished task.

Reading Skills

Reading Fluency: Word Reading

A subtest (ALLU TL2: version A) of the nationally standardized reading test battery (Lindeman, 1998) was used to assess word-level reading fluency and accuracy. Each item contained a picture with four words next to it. The task was to read the four phonologically similar words as fast as possible and draw a line connecting each picture to the word that matched it within the 2-minute time limit. The number of correct items within the time limit was used as the score representing the children's word reading fluency (maximum 80). The numbers of attempted, incorrect and missing answers (items which a child had skipped) were coded as separate variables in the data.

The task material Ied a front page and eight test sheets. The front page had four practice items, and each of the test sheets had five rows of which each had two items.

The children were asked to read words as fast as possible and to draw a line connecting each picture to the correct word. The children were encouraged to guess the correct answer in case they could not yet read the words. First, the children were asked to complete the practice items. The practice items were quickly checked together with the class prior to giving them permission to continue to the actual task. In the actual task the children were asked to complete as many items as fast and accurately as possible in the given time. The children were told to proceed one row at a time from left to right, and to turn the next page immediately if they completed the previous page. They were instructed to move on to the following item if they found some of them too challenging. They were also instructed to correct a mistake if they made one by smudging the incorrect line and drawing a new line from the picture to the correct word.

Reference:

Lindeman, J. (1998). *ALLU—Ala-asteen lukutesti* [Reading test for primary school]. University of Turku.

Reading Fluency: Sentence Reading

The TOSREC test of silent reading efficiency and comprehension was used as a measure of sentence-level reading fluency and accuracy (Wagner et al., 2009). In the test children read and evaluated the truthfulness of sentences based on real-world knowledge (e.g., "You can look at

the book" and "An elephant is small"). The task was to verify the truthfulness of as many sentences as possible out of 60 items within the 3-minute time limit. The number of correct items within the time limit was used as the score representing the children's sentence reading fluency (maximum 60). The numbers of attempted, incorrect and missing answers (items which a child had skipped) were coded as separate variables in the data.

The task material included a front page and three test sheets. The front page had two examples and five practice items, and each of the test sheets had 20 items that were placed one below another.

The children were asked to read each sentence fast and carefully and think whether the sentence is true or not. They were instructed to circle the word "Yes," if the sentence was true, and to circle "No," if the sentence was false. Sentences were constructed in a way that the truthfulness of the sentence would be easy to decide, and therefore, the variation in scores between individuals would mainly reflect their differences in reading fluency. First, the children were asked to complete the practice items. The practice items were quickly checked together with the class prior to giving them permission to continue to the actual task. In the actual task the children were asked to complete as many items as accurately as possible in the given time. The children were told to proceed one sheet at a time from top to bottom and to turn the page immediately if they completed the previous page. They were instructed to move on to the following item if they found some of them too challenging. They were also instructed to correct a mistake if they made one by smudging the incorrect answer and circling the correct word.

Reference:

Wagner, R. K., Torgesen, J. K., Rashotte, C. A., & Pearson, N. A. (2009). *TOSREC: Test of sentence reading efficiency and comprehension.* Pro-Ed.

Domain-General Cognitive Skills

Motor Speed (Copying Numbers)

Motor speed was assessed using a number copying task (Koponen, 2020). The task was to copy a given number (between 1 and 10) as fast and accurately as possible. The task included 90 items and had a 45-second time limit. The number of correct items within the time limit was used as the children's score representing motor speed (maximum 90 points). The numbers of attempted, incorrect and missing answers (items which a child had skipped) were coded as separate variables in the data.

The task Included a front page and two test sheets. The front page had two examples and two practice items, and the test sheets had 45 items arranged in three columns. Each column had 15 items.

The children were asked to copy the given number right next to each number. First, the children were asked to complete the practice items. The practice items were quickly checked together with the class prior to giving them permission to continue to the actual task. In the actual task the children were asked to complete as many items as fast and accurately as possible in the given time. The children were told to proceed one column at a time from top to bottom and after completing a column to continue to the following one. They were also instructed to turn the page immediately if they completed the first page. In addition, the children were instructed

to correct a mistake if they made one by smudging the incorrect number and drawing the correct one next to it.

Reference:

Koponen, T. (2020). *Motor speed-copying digits.* Unpublished task.

5.1.6. CA – School Phase: Grade 3, Individual Assessment (T7)

Verbal Counting Fluency

Verbal counting fluency was assessed using four different subtasks (Koponen & Salminen, 2016). The task was to count aloud numbers in the specified way (e.g., starting from a specific number, or counting forwards/backwards), as fast and as accurately as possible. In three subtasks, a 30-second time limit was used, and in one subtask the speed was measured as the time used for the subtask. Correctly recited numbers per second was used as a score for each subtask and the mean of standardized sub-scores as a score for whole task. Variables in each subtask in the data were as follows: the last recited number, incorrectly recited numbers, missing numbers (a child skipped an item) and possible extra recited numbers.

The task included two sheets, of which each had two subtasks. Each of the subtasks had both a practice item and a test item.

In the first subtask, the child was asked to count aloud numbers forwards, starting from 17 (i.e., 17, 18, 19...). The task was to count as many numbers as possible in 30 seconds. In the second subtask the child was asked to count aloud every other number forwards, starting from 1 (i.e., 1, 3, 5...). The task was to count as many numbers as possible in 30 seconds. In the third subtask the child was asked to count numbers backwards, starting from 20 and finishing at 0 (i.e., 20, 19, 18…0). The task was to count numbers as fast and accurately as possible, and the speed was measured as the time used for the subtask. In the fourth subtask the child was asked to count aloud numbers backwards, starting from 52 (i.e., 52, 51, 50...). The task was to count as many numbers as possible in 30 seconds.

Reference:

Koponen, T., & Salminen, J. (2016). *Verbal counting test*. Unpublished task.

Magnitude Comparison

The magnitude comparison skill was assessed using a magnitude comparison fluency task (Nosworthy et al., 2013). The task was to identify which of the two boxes presented included more dots. Each of the boxes had 2 to 9 dots. A maximum of 56 items could be attempted within a 1-minute time limit. One point was given from each correct answer and the sum of correct answers within the time limit was used as the score for the task (maximum 56 points). The numbers of attempted, incorrect and missing answers (items which a child had skipped) were coded as separate variables in the data.

The task included a front page and four test sheets. The front page had three examples and nine practice items (each containing two boxes with different numbers of dots varying in size, as

well). Each of the test sheets had altogether 14 items that were placed in five rows (four rows contained three items and one row contained two items).

The child was asked to look at the two adjacent boxes with dots in them and to estimate the amount of the dots instead of counting them. They were asked to cross out the box that included more dots. First, the child was asked to complete the practice items. The practice items were quickly checked together with the child prior to giving them permission to continue to the actual task. In the actual task the child was asked to complete as many items as fast and accurately as possible in the given time. The child was instructed to proceed one row at a time, from left to right, and to turn the page immediately if they completed the first page. They were also instructed to correct a mistake if they made one by smudging the incorrect line and checking the correct box.

Reference:

Nosworthy, N., Bugden. S., Archibald. L., Evans. B., & Ansari, D. (2013). A two-minute paper-and-pencil test of symbolic and nonsymbolic numerical magnitude processing explains variability in primary school children's arithmetic competence. *PloS ONE 8*(7): e67918. https://doi.org/10.1371/journal.pone.0067918

RAN (Rapid Automatized Naming)

Rapid automatized naming (see Denckla & Rudel, 1976; Heikkilä et al., in preparation) was assessed with four subtests (objects, digits, letters, and quantities). The child's task was to name the given items as fast and accurately as possible. Each subtest was scored separately. Time in seconds to name all presented items within each subtest was used as the score for each task. The number of incorrectly named items with each subtest was coded as a variable in the data.

The first three subtests (objects, digits and letters) had a series of five different visual stimuli, and the last subtest (quantities) had a series of four different visual stimuli. The stimuli for objects were pictures of cars, houses, fishes, pencils and balls; for digits the stimuli were the numbers 1, 3, 4, 5 and 6; for letters A, O, P, S and T; and for quantities, 1 to 4 dots placed randomly in a box. In the first three subtests the items were presented on a sheet in a matrix of 50 items arranged in five rows of 10 items (Ahonen et al.,1999). In the last subtest (amounts), the items were presented on a sheet in a matrix of 40 items arranged in four rows of 10 items. In each subtest the stimuli were presented in varying and random order to make sure the same stimuli would not appear sequentially.

The child was asked to name every item on the paper one-by-one as fast and accurately as they could. They were also asked to correct a mistake if they made one.

References:

Ahonen, T., Tuovinen, S., & Leppäsaari, T. (1999). *Nopean sarjallisen nimeämisen testi* [The test of rapid serial naming]. Niilo Mäki Instituutti & Haukkarannan koulu.

Denckla, M. B., & Rudel, R. G. (1976). Rapid 'automatized' naming (RAN): Dyslexia differentiated from other learning disabilities. *Neuropsychologia*, *14*(4), 471–479.

Heikkilä, R., Huotari, S., Salmi, P., Korpivaara, P., Karhunen, R., Torppa, M., Aro, M., & Ahonen, T. *Nopean nimeämisen testi – RAN* [Test of rapid automized naming – RAN] (measure in preparation). Niilo Mäki Instituutti.

Verbal Short-Term Memory and Working Memory: Word Span

Verbal short-term memory and working memory were assessed using the word span task (Koponen & Aro, 2016). The researchers presented orally a series of two-syllable words with increasing length (from two to eight words) at a rate of one word per second. In the verbal short-term memory task, the child's task was to recall and repeat the words in the presented order, whereas in the working memory task, the child's task was to repeat the words in the reversed order. Both tasks included 14 items. One point was given from each correct answer and the sum of correct answers separately for each task was used as the score for verbal short-term memory and working memory, respectively (maximum 14).

Both tasks had seven levels of difficulty (including two items with the same length, 2 to 8 words), and two different trials were presented at each level. If child failed in both items at a certain level, the task was terminated. A maximum of 14 points were given for each task.

Reference:

Koponen, T., & Aro, M. (2016). *Word span task.* Unpublished test material.

Visuo-Spatial Short-Term Memory and Working Memory: Corsi Blocks

Visuo-spatial short-term memory and working memory were assessed using the Corsi Blocks task (Corsi, 1972, see also Kessels et al., 2000, 2008). The task was to memorize the sequence in which the researcher tapped blocks arranged in front of the child. The researcher tapped the blocks at a rate of one tap per second. In the visuo-spatial short-term memory task, the child's task was to tap the blocks in the presented order, whereas in the working memory task, the child's task was to tap the blocks in the reversed order. The visuo-spatial short-term memory task included 16 items and the working memory task 14 items. One point was given from each correct response and the sum of correct responses separately for each task was used as the score for visuo-spatial short-term memory and working memory, respectively (maximum 16 and 14, respectively).

In the task, nine blocks (cubes, 3 x 3 cm) were attached asymmetrically to a board that was placed on a table in between the researcher and the child. Each block had a number from one to nine on one side, and only the researcher could see the numbers.

The task included two subtasks. In the visuo-spatial short-term memory task the child was asked to tap the blocks in the same order as the researcher did. The task included altogether 16 items, two items in eight levels of difficulty. The amount of taps to be repeated increased by one for each level – starting with two taps on the first level and ending with nine taps on the last level. In the working memory task the child was asked to tap the blocks in reverse order compared to the researcher. The task had seven levels of which each had two items, meaning there were altogether 14 items. The amount of taps to be repeated increased by one for each level – starting with two taps in the first level and ending with eight taps in the last level. Each task was continued until the child made a mistake in both items on the same level.

References:

Corsi, P.M. (1972). *Human memory and the medial temporal region of the brain* [Doctoral Dissertation, McGill University]. Retrieved from https://escholarship.mcgill.ca/concern/theses/05741s554

Kessels, R. P., van Den Berg, E., Ruis, C., & Brands, A. M. (2008). The backward span of the Corsi Block-Tapping Task and its association with the WAIS-III Digit Span. *Assessment, 15*(4), 426–434.

Kessels, R.P.C., van Zandvoort, M.J.E., Postman, A., Kapelle, L.J., & de Hand, E.H.F. (2000). The Corsi Block-Tapping Task: Standardization and Normative Data. *Applied Neuropsychology, 7(4)*, 252–258.

Spatial Relations

Spatial perception was evaluated with the Spatial relations test (Woodcock & Johnson, 1977, see also Schrank, 2006). The task assesses the child's skill to rotate and process spatial forms mentally. A maximum of 31 items could be attempted within a 3-minute time limit. One point was given from each correct answer and the sum of correct answers within the time limit was used as the score for the task (maximum 31). The numbers of attempted, incorrect and missing answers (items which a child had skipped) were coded as separate variables in the data.

The task included altogether 15 pages: a front page with an example item, followed by two pages with one practice item per each page, and finally 12 test sheets. The first and the last test sheets consisted of one item, the second test sheet consisted of two items, one below another, and the rest of the test sheets consisted of three items, one below another.

The child was shown colorful stimulus pictures of a pattern formed by different pieces on the left side of the paper. On the right side there were six different pieces that could come together to form the pattern on the left. The child was asked to decide which pieces (from two to three) in the stimulus picture formed the pattern if combined, and to circle the letters under those pieces. First, the child was asked to complete the practice items. The practice items were quickly checked together with the child prior to giving them permission to continue to the actual task. In the actual task the child was asked to complete as many items as fast and accurately as possible in the given time. The child was asked to proceed one item at a time, and to turn the page immediately if they completed all the items on a page. They were also instructed to correct a mistake if they made one by smudging the incorrect circle and to circle the correct letters.

References:

Schrank, F. A. (2006). Specification of the cognitive processes involved in performance on the Woodcock-Johnson III. *Assessment Service Bulletin, (7).* Riverside Publishing.

Woodcock, R. W., & Johnson, M. B. (1977). *Woodcock-Johnson Psycho-Educational Battery*. Riverside Publishing.

5.2. CR – Children's Reports

5.2.1. CR – Preschool Phase (T3–T4)

Children's Interest in Letters/Reading and Numbers/Math

Child's interest in tasks related to letters and reading and in tasks related to numbers and math was assessed through interviews (modified from Aunola & Nurmi, 1999 and Nurmi & Aunola,

2005). The interview included three items related to each content area. In each item, the child responded by selecting an appropriate option on a 5-point Likert scale. The average of the three responses to questions related to each content area was used as the score representing the child's interest in literacy or numeracy tasks.

The researcher asked the questions and the child responded by pointing to the best-suited option from five smiley faces (Likert scale with facial expressions varying from an unhappy / frowning face [1 = *very boring / I do not like doing this at all*] to very smiley face [5 = *very nice / I like doing this very much*]). The child was asked three questions related to letters and reading and three questions related to numbers and counting. Two questions were asked in the context of ECEC and the third question in the context of home. For example, "How glad are you while doing tasks including letters and reading / tasks including numbers and counting in ECEC?" or "How much do you like doing tasks including letters and reading / tasks including numbers and counting at home?"

References:

Aunola, K. & Nurmi, J.-E. (1999). *Task-value scale for children.* Unpublished task.

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.* https://doi.org/10.1016/j.learninstruc.2005.04.009

5.2.2. CR – School Phase: Grade 2 (T6)

Children's Interest in Reading and Math

Children's interest in reading and math was assessed through self-evaluation questions (modified from the task-value scale in Aunola & Nurmi, 1999 and Nurmi & Aunola, 2005). The questionnaire included six items of which three related to reading (e.g., "How much do you like reading tasks at school?") and three related to math (e.g., "How much do you like math tasks at school?").

The questionnaire Included one sheet that had six items, one below another. Five smiley faces were placed next to each question. The faces represented the feelings towards reading/math.

The questionnaires were conducted in a classroom where the researcher asked the questions out loud and the children had their individual response sheets in front of them. Children were instructed to think about how much they enjoyed reading and math tasks at home and in school. The researcher read aloud all six items, one at a time, and the children were asked to mark their answers on their own assessment sheet by selecting the best-suited option from five smiley faces (Likert scale with facial expressions varying from frowning face [1 = *very boring / I do not like doing this at all*] to very smiley face [5 = *very nice / I like doing this very much*]).

References:

Aunola, K. & Nurmi, J.-E. (1999). *Task-value scale for children*. Unpublished task.

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.* https://doi.org/10.1016/j.learninstruc.2005.04.009

Children's Reading and Math Anxiety

Children's anxiety toward reading and math was assessed through self-evaluation questions (Sorvo et al., 2017, 2019) in which children were instructed to think about how they felt when they read or did math tasks. The questionnaire included six items, of which three concerned reading anxiety (e.g., "I feel anxious when I have to respond in the math lesson") and three items concerned math anxiety (e.g., "I feel anxious when I know that I have to read aloud in the lesson").

The questionnaire included one sheet that had six items, one below another. Empty answer boxes were placed next to each question.

The questionnaires were conducted in a classroom where the researcher asked the questions out loud and the children had their individual response sheets in front of them. Before reading aloud the items, the researcher showed children a picture of two children who were seemingly anxious (anxious facial expression and hand lifted close to face) and read aloud the text below the picture saying: "Sometimes the learning situations might make you feel anxious. Anxiety might, for example, feel like not being able to think clearly or that – 'I don't remember or I can't do anything!'. Sometimes one's stomach or head might hurt or you might feel the anxiety within your entire body." The researcher read aloud all six items, one at a time, and children were asked to mark their answers on their own assessment sheet by selecting the most well-suited alternative on a 5-point Likert scale (1 = *I do not feel like this at all*, 5 = *I feel like this very much*).

References:

Sorvo, R., Koponen, T., Viholainen, H., Aro, T., Räikkönen, E., Peura, P., Dowker, A., & Aro, M. (2017). Math anxiety and its relationship with basic arithmetic skills among primary school children. *British Journal of Educational Psychology, 87*(3), 309–327. https://doi.org/10.1111/bjep.12151

Sorvo, R., Koponen, T., Viholainen, H., Aro, T., Räikkönen, E., Peura, P., Tolvanen, A., & Aro, M. (2019). Development of math anxiety and its longitudinal relationships with arithmetic achievement among primary school children. *Learning and Individual Differences, 69*, 173–181. https://doi.org/10.1016/j.lindif.2018.12.005

5.2.3. CR – School Phase: Grade 3 (T7)

Children's Interest in Reading and Math

Children's interest in reading and math was assessed through self-evaluation questions (modified from the task-value scale in Aunola & Nurmi, 1999 and Nurmi & Aunola, 2005). The questionnaire included six items of which three related to reading (e.g., "How much do you like reading tasks at school?") and three related to math (e.g., "How much do you like math tasks at school?").

The questionnaire Included one sheet that had six items, one below another. Five smiley faces were placed next to each question. The faces represented the feelings towards reading/math.

The questionnaires were conducted in a classroom where the researcher asked the questions out loud and the children had their individual response sheets in front of them. Children were

instructed to think about how much they enjoyed reading and math tasks at home and in school. The researcher read aloud all six items, one at a time, and the children were asked to mark their answers on their own assessment sheet by selecting the best-suited option from five smiley faces (Likert scale with facial expressions varying from frowning face [1 = *very boring/ I do not like doing this at all*] to very smiley face [5 = *very nice / I like doing this very much]).*

References:

Aunola, K. & Nurmi, J.-E. (1999). *Task-value scale for children*. Unpublished task.

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.* https://doi.org/10.1016/j.learninstruc.2005.04.009

Children's Reading and Math Anxiety

Children's anxiety toward reading and math was assessed through self-evaluation questions (Sorvo et al., 2017, 2019) in which children were instructed to think about how they felt when they read or did math tasks. The questionnaire included six items, of which three concerned reading anxiety (e.g., "I feel anxious when I have to respond in the math lesson") and three items concerned math anxiety (e.g., "I feel anxious when I know that I have to read aloud in the lesson").

The questionnaire included one sheet that had six items, one below another. Empty answer boxes were placed next to each question.

The questionnaires were conducted in a classroom where the researcher asked the questions out loud and the children had their individual response sheets in front of them. Before reading aloud the items, the researcher showed children a picture of two children who were seemingly anxious (anxious facial expression and hand lifted close to face) and read aloud the text below the picture saying: "Sometimes the learning situations might make you feel anxious. Anxiety might, for example, feel like not being able to think clearly or that – 'I don't remember or I can't do anything!'. Sometimes one's stomach or head might hurt or you might feel the anxiety within your entire body." The researcher read aloud all six items, one at a time, and children were asked to mark their answers on their own assessment sheet by selecting the most well-suited alternative on a 5-point Likert scale (1 = *I do not feel like this at all*, 5 = *I feel like this very much*).

References:

Sorvo, R., Koponen, T., Viholainen, H., Aro, T., Räikkönen, E., Peura, P., Dowker, A., & Aro, M. (2017). Math anxiety and its relationship with basic arithmetic skills among primary school children. *British Journal of Educational Psychology*, *87*(3), 309–327. https://doi.org/10.1111/bjep.12151

Sorvo, R., Koponen, T., Viholainen, H., Aro, T., Räikkönen, E., Peura, P., Tolvanen, A., & Aro, M. (2019). Development of math anxiety and its longitudinal relationships with arithmetic achievement among primary school children. *Learning and Individual Differences, 69*, 173–181. https://doi.org/10.1016/j.lindif.2018.12.005

5.3. RR – Research Assistants' Reports Concerning Individual Child

5.3.1. RR – Preschool Phase (T3–T4)

Task Behavior

Research assistants rated children's task behavior in the assessment situations in the fall 2018 and in the spring 2019 using the Observer-Rating Scale of Achievement Strategies (OSAS; Nurmi & Aunola, 1998; for more information on the validity of the scales in the Finnish samples, see Zhang et al., 2011). After testing each child on an individual basis, the research assistants rated on a 7-point scale how well the 10 items presented described the child's behavior during the assessment situation (1 = *not at all this kind of behavior,* 4 = *this kind of behavior is somewhat typical,* 7 = *always or almost all the time this kind of behavior*).

Task behavior comprised five domains, each rated with two items: *Active task avoidance* (two items: "If there are problems with the task, the child starts doing something else" and "If there are problems with the task, the child becomes interested in other things in the room"), *Passive task avoidance* (two items: "If there are problems with the task, the child stops doing it and waits passively" and "The child gives up easily [or says, they are not able to the tasks already before they start]"), *Task focus* (two items: "Although the task turns difficult for the child, the child tries hard to finish it" and "The child tries persistently to do the tasks [even though it is hard, or the tasks are difficult]"), *Social dependence* (two items: "If there are problems with the task, the child seeks for your support" and "The child seeks for your support when doing the task") and *Anxiety* (two items: "If there are problems with the task, the child becomes anxious" and "The child is afraid of difficult tasks.")

References:

Nurmi, J.-E., & Aunola, K. (1998). *Observer Rating Scale of Achievement Strategies (OSAS).* Unpublished measurement instrument.

Zhang, X., Nurmi, J.-E., Kiuru, N., Lerkkanen, M.-K., & Aunola, K. (2011). A teacher-report measure of children's task-avoidant behavior: A validation study of the behavioral strategy rating scale. *Learning and Individual Differences, 2,* 690–698. https://doi.org/10.1016/j.lindif.2011.09.007

5.3.2. RR – School Phase: Grade 3, Individual Assessment (T7)

Task Behavior

Research assistants rated children's task behavior in the individual assessment situations in the spring 2023 using the Observer-Rating Scale of Achievement Strategies (OSAS; Nurmi & Aunola, 1998; for more information on the validity of the scales in the Finnish samples, see Zhang et al., 2011). After testing the child on an individual basis, the research assistants rated on a 7-point scale how well the 10 items presented described the child's behavior during the individual assessment situation (1 = *not at all this kind of behavior,* 4 = *this kind of behavior is somewhat typical,* 7 = *always or almost all the time this kind of behavior*).

Task behavior comprised five domains, each rated with two items: *Active task avoidance* (two items: "If there are problems with the task, the child starts doing something else" and "If there are problems with the task, the child becomes interested in other things in the room"), *Passive task avoidance* (two items: "If there are problems with the task, the child stops doing it and waits passively" and "The child gives up easily [or says, they are not able to the tasks already before they start]"), *Task focus* (two items: "Although the task turns difficult for the child, the child tries hard to finish it" and "The child tries persistently to do the tasks [even though it is hard, or the tasks are difficult]"), *Social dependence* (two items: "If there are problems with the task, the child seeks for your support" and "The child seeks for your support when doing the task") and *Anxiety* (two items: "If there are problems with the task, the child becomes anxious" and "The child is afraid of difficult tasks.")

References:

Nurmi, J.-E., & Aunola, K. (1998). *Observer Rating Scale of Achievement Strategies (OSAS).* Unpublished measurement instrument.

Zhang, X., Nurmi, J.-E., Kiuru, N., Lerkkanen, M.-K., & Aunola, K. (2011). A teacher-report measure of children's task-avoidant behavior: A validation study of the behavioral strategy rating scale. *Learning and Individual Differences, 2*, 690–698. https://doi.org/10.1016/j.lindif.2011.09.007

5.4. PA – Parental Assessments

Reading Fluency and Accuracy: Word List Reading & Pseudoword List Reading

List reading fluency was assessed with two lists, one with words and another with pseudowords (Nevala et al., 2006). There were 30 items in each of the lists. The reading speed score for each of the lists was the time used to read all 30 items. The reading accuracy score for each of the lists was the amount of items read correctly (maximum 30 points).

In each of the tasks, the parent was asked to read the list aloud as fast and as accurately as they could. The word list included common Finnish words with five to 16 letters and with two to five syllables. The pseudowords were similar in length and structure to the real word items and were created by exchanging consonants and vowels of the items of the word list.

Reference:

Nevala, J., Kairaluoma, L., Ahonen, T., Aro, M., & Holopainen, L. (2006). *Lukemis- ja kirjoittamistaitojen yksilötestistö nuorille ja aikuisille* [Individual reading and writing test battery for young people and adults]. Niilo Mäki Institute.

Reading Fluency and Accuracy: Text Reading

In the text reading task (Leinonen et al., 2001), the parent was asked to read a text on Lapland aloud as fast and accurately as possible. The text consisted of 211 words. The score for text reading speed was the time used to read the text and the accuracy score was the amount of correctly read words.

Reference:

Leinonen, S., Müller, K., Leppänen, P. H., Aro, M., Ahonen, T., & Lyytinen, H. (2001). Heterogeneity in adult dyslexic readers: Relating processing skills to the speed and accuracy of oral text reading. *Reading and Writing*, *14*(3–4), 265–296. https://doi.org/10.1023/A:1011117620895

RAN (Rapid Automatized Naming)

Rapid automatized naming (see Denckla & Rudel, 1976) was assessed with three subtests (objects, digits and letters). In RAN, the task is to name the given items as fast and accurately as possible. Each subtest was scored separately. Time in seconds to name all presented items within each subtest was used as the score for each task. The number of incorrectly named items with each subtest was coded as a variable in the data.

Each subtest had a series of five different visual stimuli. The stimuli for objects were pictures of cars, houses, fishes, pencils and balls; for digits, the stimuli were the numbers 2, 4, 6, 7 and 9; and for letters, A, O, P, S and T. The items were presented on a sheet in a matrix of 50 items arranged in five rows of 10 items (Ahonen et al., 1999). In each subtest the stimuli were presented in varying and random order to make sure the same stimuli would not appear sequentially.

The parent was asked to name every item on the paper one by one as fast and accurately as they could. They were also asked to correct a mistake if they made one.

In addition, a subsample of 49 parents completed both the older Finnish version of the RAN (Ahonen et al., 1999) and the revised version (Heikkilä et al., in preparation). The revised version included four RAN subtests (objects, digits, letters and quantities). The stimuli for objects were pictures of cars, houses, fishes, pencils and balls; for digits the stimuli were numbers 1, 3, 4, 5 and 6; for letters, A, O, P, S and T; and for quantities, 1 to 4 dots placed randomly in a box. In the first three subtests, the items were presented on a sheet in a matrix of 50 items arranged in five rows of 10 items. In the last subtest (quantities), the items were presented on a sheet in a matrix of 40 items arranged in four rows of 10 items. In each subtest the stimuli are presented in varying and random order to make sure the same stimuli would not appear sequentially. The parents were asked to perform the task in a similar way as explained previously.

References:

Ahonen, T., Tuovinen, S., & Leppäsaari, T. (1999). *Nopean sarjallisen nimeämisen testi* [The test of rapid serial naming]. Niilo Mäki Instituutti & Haukkarannan koulu.

Denckla, M. B., & Rudel, R. G. (1976). Rapid ‘automatized’ naming (RAN): Dyslexia differentiated from other learning disabilities. Neuropsychologia, 14(4), 471–479.

Heikkilä, R., Huotari, S., Salmi, P., Korpivaara, P., Karhunen, R., Torppa, M., Aro, M., & Ahonen, T. *Nopean nimeämisen testi – RAN* [Test of rapid automized naming – RAN] (measure in preparation). Niilo Mäki Instituutti.

Motor Speed (Copying Numbers)

Motor speed was assessed using a number copying task (Koponen, 2020). The task was to copy a given number (between 1 and 10) as fast and accurately as possible. The task included 45

items, and the parent was instructed to finish the whole task. The time was measured, and the correct answers per second was used as the score for representing the parent's motor speed. The numbers of attempted, incorrect and missing answers (items which a parent had skipped) were coded as separate variables in the data.

The task included a front page and one test sheet. The front page had two examples and two practice items, and the test sheets had three columns of items. Each column had 15 items.

The parent was asked to copy the given number right next to each number. First, the parent was asked to complete the practice items. The practice items were quickly checked by the research assistant prior to giving the parent permission to continue to the actual task. In the actual task the parent was asked to complete the whole task as fast and accurately as possible. The parent was instructed to proceed one column at a time, from top to bottom, and after completing a column to continue to the following one. They were also instructed to correct a mistake if they made one by smudging the incorrect number and drawing the correct one next to it.

Reference:

Koponen, T. (2020). *Motor speed-copying digits*. Unpublished task.

Magnitude Comparison

The magnitude comparison skill was assessed using a magnitude comparison fluency task adapted from the Numeracy Screener (Nosworthy et al., 2013). The task was to identify which of the two boxes presented included more dots. Each of the boxes had 2 to 9 dots. The task included 30 items, and the parent was instructed to finish the whole task as fast and accurately as possible. The time was measured, and the correct answers per second was used as the score for representing the parent's magnitude comparison skills. The numbers of attempted, incorrect and missing answers (items which a parent had skipped) were coded as separate variables in the data.

The task included a front page and two test sheets. The front page had two examples and two practice items (each containing two boxes with different numbers of dots varying in size, as well). The test sheets had three columns of which each had five items.

The parent was asked to look at the two adjacent boxes with dots in them and to estimate the amount of the dots instead of counting them. They were asked to cross out the box that included more dots. First, the parent was asked to complete the practice items. The practice items were quickly checked by the research assistant prior to giving the parent permission to continue to the actual task. In the actual task the parent was asked to complete the whole task as fast and accurately as possible. The parent was instructed to proceed one column at a time, from top to bottom, and after completing a column to continue to the following one. They were also asked to turn the page immediately after completing all items on a page. In addition, they were instructed to correct a mistake if they made one by smudging the incorrect line and checking the correct box.

Reference:

Nosworthy, N., Bugden, S., Archibald, L., Evans, B., & Ansari, D. (2013). A two-minute paper-and-pencil test of symbolic and nonsymbolic numerical magnitude processing

explains variability in primary school children's arithmetic competence. *PloS ONE 8*(7): e67918. https://doi.org/10.1371/journal.pone.0067918

Number Comparison

Number comparison skill was assessed using the paper and pencil task where the task was to identify the larger of the two single-digit numbers (from one to nine) (Brankaer et al., 2017). The task included 60 items, and the parent was instructed to finish the whole task as fast and accurately as possible. The time was measured, and the correct answers per second was used as the score for representing the parent's number comparison skills. The numbers of attempted, incorrect and missing answers (items which a parent had skipped) were coded as separate variables in the data.

The task included a front page and one test sheet. The front page had two examples and two practice items. The test sheet had four columns of which each had 15 items (a box with two numbers).

The parent was asked to cross out the number that was greater than the adjacent number in the same box. First, the parent was asked to complete the practice items. The practice items were quickly checked by the research assistant prior to giving the parent permission to continue to the actual task. In the actual task the parent was asked to complete the whole task as fast and accurately as possible. The parent was instructed to proceed one column at a time, from top to bottom, and after completing a column to proceed to the following column. They were also instructed to correct a mistake if they made one by smudging the incorrect line and then selecting the correct number.

Reference:

Brankaer, C., Ghesquière, P., & De Smedt, B. (2017). Symbolic magnitude processing in elementary school children: A group administered paper-and-pencil measure (SYMP Test). *Behavior research methods, 49*, 1361–1373. https://doi.org/10.3758/s13428-016-0792-3

Number Relation Fluency

Number concept knowledge was assessed using the number relation task (Koponen, 2021a), where the task was to detect the exact magnitude difference between two numbers, that is, how much bigger one of the two given numbers was. The number range was between one and ten and the difference in magnitude varied from one to three. The task included 40 items, and the parent was instructed to finish the whole task as fast and accurately as possible. The time was measured, and the correct answers per second was used as the score for representing the parent's number relations skills. The numbers of attempted, incorrect and missing answers (items which a parent had skipped) were coded as separate variables in the data.

The task included a front page and one test sheet. The front page had two examples and two practice items. The test sheet had four columns, of which each had 10 items (two numbers presented next to each other and an empty answer box below them).

The parent was asked to mark how much greater one of the two numbers was in comparison with the other number in the item. First, the parent was asked to complete the practice items. The practice items were quickly checked by the research assistant prior to giving the parent

permission to continue to the actual task. In the actual task, the parent was asked to complete the whole task as fast and accurately as possible. The parent was instructed to proceed one column at a time, from top to bottom, and after completing a column to proceed to the following column. They were also instructed to correct a mistake if they made one by smudging the incorrect number and drawing the correct one next to it.

Reference:

Koponen, T. (2021a) *Number relations task*. Unpublished task.

Arithmetic Fluency

In the arithmetic fluency task (Aunola & Räsänen, 2007) the task was to calculate additions (e.g., 732 + 8 = __), subtractions (e.g., 57 – __ = 48), multiplications (e.g., 11 x 3,2 = __), divisions (e.g., 240 / 8 = __) or arithmetic problems with several operations (e.g., 40 / 8 – 3 = __) as fast and accurately as possible. A maximum of 28 items could be attempted within a 3-minute time limit. One point was given for each correct answer and the sum of correct answers was used as the score representing arithmetic fluency (maximum 28). The numbers of attempted, incorrect, and missing answers (items which a parent had skipped) were coded as separate variables in the data.

The task included a front page and two test sheets. The front page had two examples and two practice items. The test sheets had seven rows, of which each had two items.

The parent was asked to do the calculations and write the answer on the empty line. First, the parent was asked to complete the practice items. The practice items were quickly checked by the research assistant prior to giving the parent permission to continue to the actual task. In the actual task, the parent was asked to complete as many items as fast and accurately as possible in the given time. They were instructed to move on to the following item if they found some of them too challenging. The parent was told to proceed one row at a time, from left to right, and to turn the page immediately if they completed the first page. They were also instructed to correct a mistake if they made one by smudging the incorrect number and drawing the correct one next to it.

Reference:

Aunola, K., & Räsänen, P. (2007). *The Basic Arithmetic Test*. Unpublished task.

Multiplication Fluency

In the multiplication fluency task (Koponen & Mononen, 2010c), the task was to calculate as many multiplications including numbers from 1 to 9 (e.g., 1 x 3 = __; 8 x 6 = __) as possible. A maximum of 120 items could be attempted within a 2-minute time limit. One point was given from each correct answer and the sum of correct answers within the time limit was used as the score for the task (maximum 120). The numbers of attempted, incorrect and missing answers (items which a parent had skipped) were coded as separate variables in the data.

The task included a front page and two test sheets. The front page had two examples and two practice items. The two test sheets had three columns, of which each had 20 items.

The parent was asked to calculate the multiplication and write the answer next to each problem. First, the parent was asked to complete the practice items. The practice items were quickly

checked by the research assistant prior to giving the parent permission to continue to the actual task. In the actual task, the parent was asked to complete as many items as fast and accurately as possible in the given time. The parent was instructed to proceed one column at a time, from top to bottom, and after completing a column to continue to the following one. They were also instructed to turn the page immediately if they completed the first page. In addition, the parent was told to correct a mistake if they made one by smudging the incorrect number and drawing the correct one next to it.

Reference:

Koponen, T., & Mononen, R. (2010c). *The 2-minute multiplication fluency test*. Unpublished task.

A Mathematical Achievement Test (RMAT)

Mathematical skill was assessed with the RMAT (Räsänen et al., 2008). The test assesses basic mathematics skills, such as arithmetic (e.g., 4444 – 400 = __; 204 x 12 = __), fractions and decimals (e.g., 3 / 10 = __; 5 × 0,3 = __), measurement (e.g., €12.50 = __ cents), and algebra tasks (e.g., x / 20 = 8). A maximum of 56 items could be attempted within a 10-minute time limit. One point was given from each correct answer and the sum of correct answers within the time limit was used as the score of the task (maximum 56). The numbers of incorrect and missing answers (items which a parent had skipped) were coded as separate variables in the data.

The task included four sheets altogether, of which the first one was a cover sheet and the two following ones were the test sheets. The last page had empty space that could be used for notes or calculations. The first test sheet had 42 items and the last test sheet had 14 items, and also some empty space for the notes.

The parent was asked to do the calculations and write the answer right next to each calculation. They were instructed to complete as many items as accurately as possible in the given time. The parent was told to proceed in order, but they were also instructed to move on to the following item if they found some of them too challenging. In addition, they were instructed to use the empty space of the paper if they needed to do calculations to reach the conclusion.

Reference:

Räsänen, P., Linnanmäki, K., Haapamäki, C., & Skagersten, D. (2008). *RMAT-Test av räknefärdighet hos evelver I åldern 9–12 år* [A mathematical achievement test for ages 9–12 in Finnish-Swedish]. Niilo Mäki Institute.

KTLT-A

Mathematical skill was assessed with the Test for Basic Mathematical Skills for Grades 7–9 KTLT-A (Räsänen & Leino, 2005), which is designed to assess mathematical skills that are usually learned by the age of 16 in Finland. The task requires basic arithmetic skills (addition, subtraction, multiplication, and division), as well as solving word problems, algebra, geometry, and unit conversion skills. This test is also used in Finland as a screening tool to identify students at risk for mathematics difficulties. A maximum of 40 items could be attempted within a 30-minute time limit. One point was given from each correct answer and the sum of correct answers within the time limit was used as the score of the task (maximum 40). The numbers of

incorrect and missing answers (items which a parent had skipped) were coded as separate variables in the data.

The task included altogether four test sheets, of which the first had 11 items, the second had 11 items, the third had 13 items, and the last page had 5 items.

The parent was asked to do the calculations and write the answer right next to each calculation. They were instructed to complete as many items as accurately as possible in the given time. The parent was asked to proceed in order, but they were also instructed to move on to the following item if they found some of them too challenging and told to return to the challenging items in case there was time left over. They were also given a blank sheet of paper in order to make some notes or to do calculations to reach the conclusion.

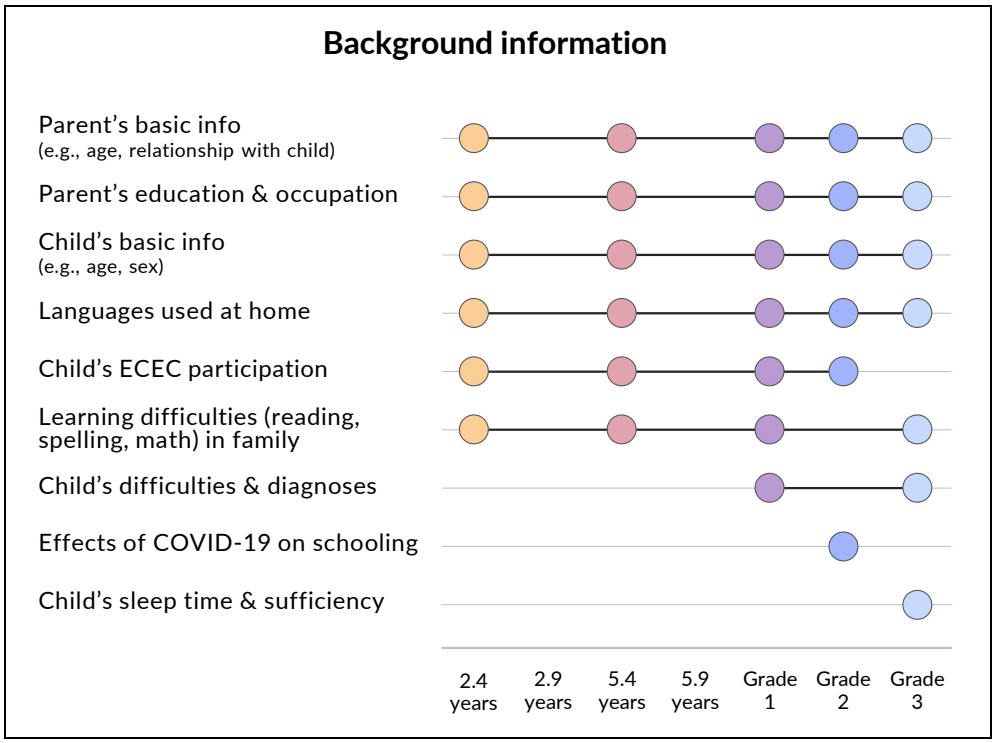
Reference:

Räsänen, P., & Leino, L. (2005). *KTLT-laskutaidon testi* [KTLT-a test of basic mathematical skills]. Niilo Mäki instituutti.

5.5. PQ – Parental Questionnaires

Parental questionnaires were sent in T1 (2.4 years), T3 (5.4 years), T5 (Grade 1), T6 (Grade 2) and T7 (Grade 3). Figures 7–11 provide an overview of the measures grouped in five domains: 1. Background information, 2. Child's reading and math motivation and temperament, 3. Home numeracy and literacy environment (HNE & HLE), 4. Child's homework and parental involvement, and 5. Parent's own reading and math skills and experiences. The measures are described in detail below the figures.

FIGURE 7 An overview of the **background information**.



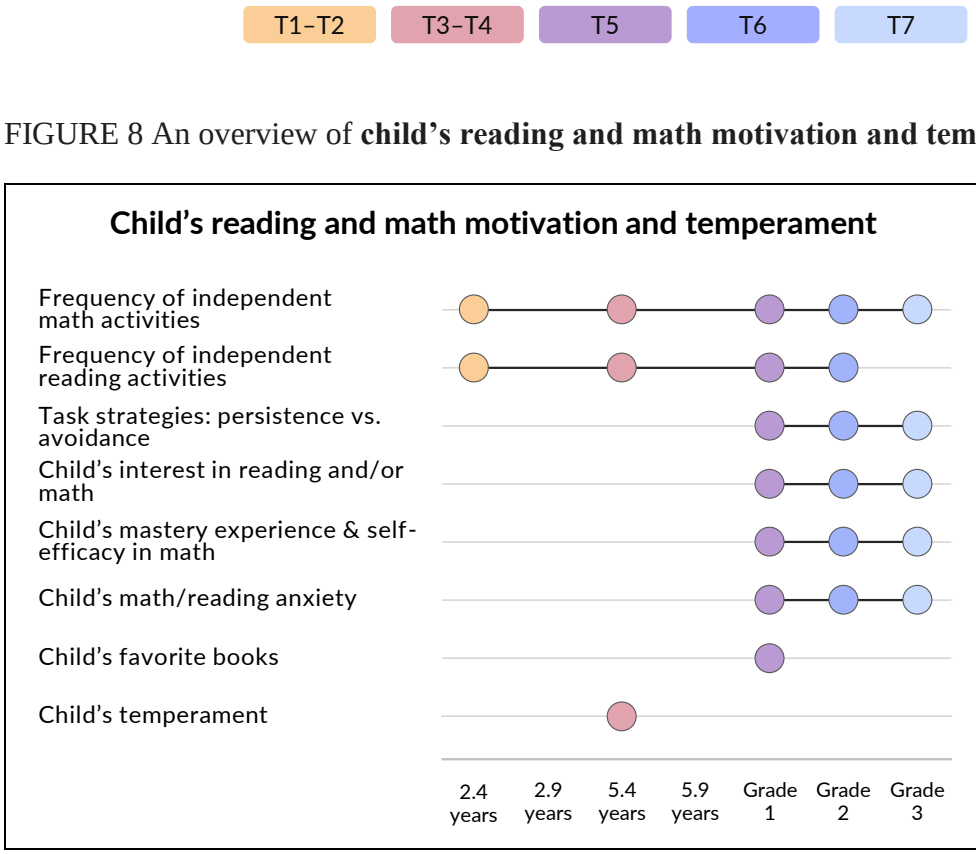


FIGURE 8 An overview of **child's reading and math motivation and temperament.**

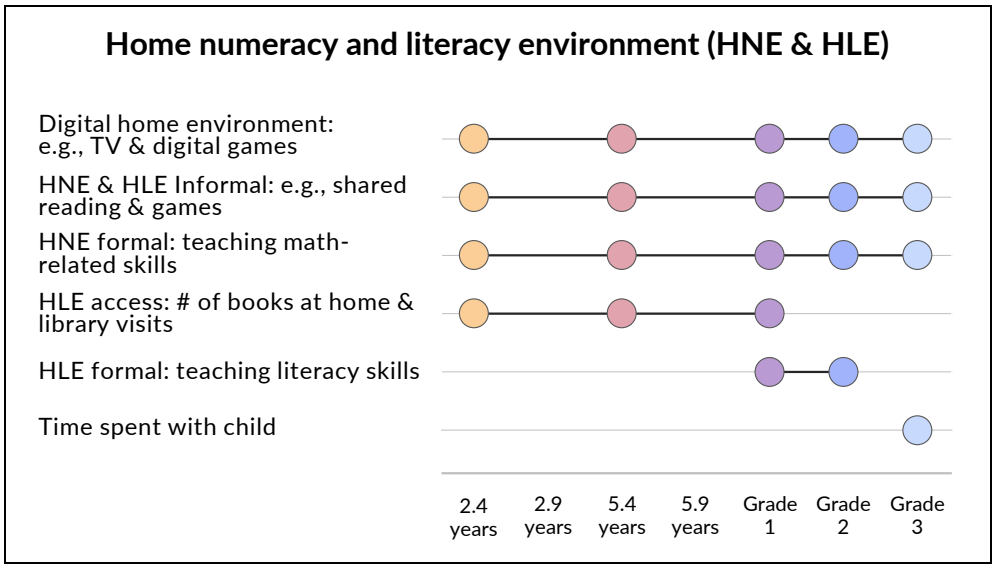


FIGURE 9 An overview of **home numeracy and literacy environment (HNE & HLE).**

FIGURE 10 An overview of **child's homework and parental involvement.**

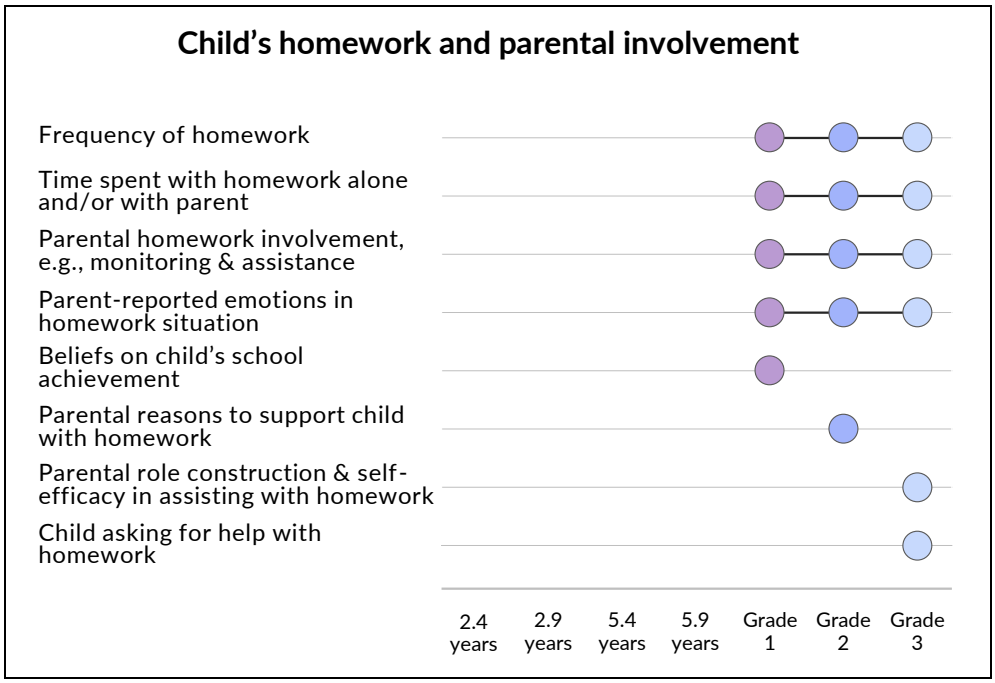
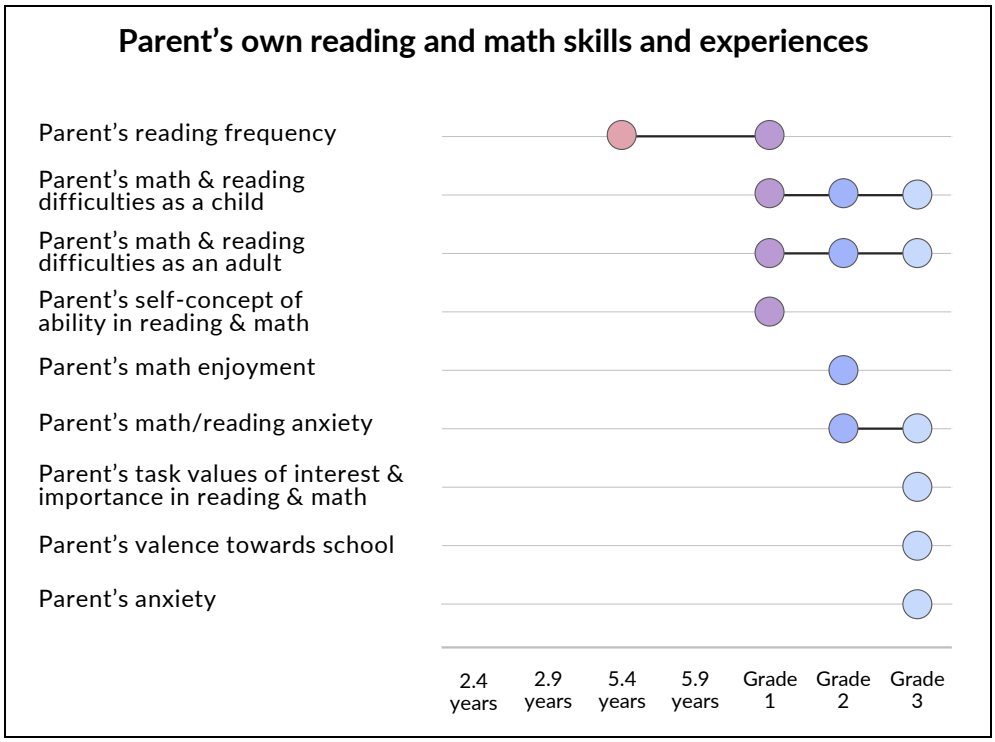


FIGURE 11 An overview of the questions regarding **parent's own reading and math skills and experiences**.



5.5.1. PQ – Toddler Phase (T1)

BACKGROUND INFORMATION

- Who filled in the questionnaire: (1) Mother; (2) Father; (3) Someone else, who: ____
- Respondent's year of birth: ____
- Living arrangements
  - Do you live (1) Together with a spouse (married) and children, (2) Together with a spouse and children; (3) In a blended family with children; (4) As a single parent with children; (5) As some other family type, what: ____
  - Do you live in the same household with the participating child: (1) Yes; (2) No; (3) Partly: ____days/month
- Parental education:
  - Own basic education: (1) Less than comprehensive school; (2) Comprehensive school; (3) General upper secondary school
  - Own vocational education: (1) No vocational education; (2) Vocational upper secondary school; (3) Vocational college education (in Finnish "opisto"); (4) Polytechnic university; (5) University
  - Basic education of child's other guardian: (1) Less than comprehensive school; (2) Comprehensive school; (3) General upper secondary school
  - Vocational education of child's other guardian: (1) No vocational education; (2) Vocational school; (3) College education (in Finnish "opisto"); (4) Polytechnic university; (5) University
- Parental employment and occupation:
  - Employment situation: (1) Working; (2) Unemployed / partly unemployed, how long to date ____; (3) Studying; (4) Suspended without pay (laid off), how long to date: ____; (5) Parental leave, how long to date ____; (6) Stay-home mother/father
  - What is your occupation/profession: ____
  - What is the occupation/profession of the child's other guardian: ____
- Participating child's birthday (day, month, year): ____
- Participating child's gender: (1) Boy; (2) Girl
- Which languages does the participating child speak at home: (1) Finnish; (2) Other, which: ____
- ECEC participation: ____hours/day, ____days/week
- What is the format of ECEC: (1) Fulltime; (2) Parttime
- The ECEC history of the child: Use the table to indicate your child's ECEC history in a chronological order (use the first row as an example). Write down the child's age and tick the correct box for other characteristics.

A. Child age	B. ECEC location			C. ECEC format			D. Duration h/day			E. Duration in months
	Center based ECEC	Home based ECEC (Family day carer)	Some-where else	half a day	full day	some other format	less than 4h	5–8 h	more than 9h	duration in months
Example: 1 y 2 mo – 2 y 3 mo		x			x			x		13
1.										
2.										
3.										

Parent-Reported Learning Difficulties in Reading and Mathematics

Parents were asked if they and/or the other parent of the child had experienced learning difficulties (1) in reading or writing, and (2) in mathematics or calculation. The questionnaire included one question about their own difficulties in reading or writing, one about their own difficulties in mathematics or calculation, and the same two items concerning the other parent. The parents answered each question on a 3-point Likert scale (1 = *no difficulties*, 2 = *some difficulties*, 3 = *clear difficulties*). The items have been used in previous studies (Lerkkanen et al., 2006–2016; Lohvansuu et al., 2021).

References:

Lerkkanen, M.-K., Niemi, P., Poikkeus, A.-M., Poskiparta, E., Siekkinen, M., & Nurmi, J.-E. (2006–2016). *The First Steps Study* [Alkuportaat]. University of Jyväskylä.

Lohvansuu, K., Torppa, M., Ahonen, T., Eklund, K., Hämäläinen, J. A., Leppänen, P. H., & Lyytinen, H. (2021). Unveiling the mysteries of dyslexia—Lessons learned from the prospective Jyväskylä longitudinal study of dyslexia. *Brain Sciences*, *11*(4), 427. https://doi.org/10.3390/brainsci11040427

Children's Reading and Math Motivation

The Frequency of Children's Math Activities

There were four items of the frequency of children's math activities adapted from Blevins-Knabe & Musun-Miller (1996). Parents were asked to respond to the following question: "In the past week, how often have the following activities happened at your home?" using a Likert scale (1 = *not at all*, 4 = *six times a week or more often).* Item examples: "Child mentioned a fact such as 1 + 1 = 2 or 4 – 2 = 2"; "Child recites the number line 1,2,3,4…".

Reference:

Blevins-Knabe, B., & Musun-Miller, L. (1996). Number use at home by children and their parents and its relationship to early mathematical performance. *Early Development and Parenting: An International Journal of Research and Practice*, *5*(1), 35–45. https://doi.org/10.1002/(SICI)1099-0917(199603)5:1<35::AID-EDP113>3.0.CO;2-0

The Frequency of Children's Literacy Activities

There were three items of the frequency of children's independent literacy activities. Parents were asked to respond to the following question: "How often do the following activities happen at your home?" using a Likert scale (1 = *not at all or rarely*, 5 = *several times a day*). Item examples: "Child 'reads' / looks at / browses through books/magazines/comic books on their own"; "Child asks to be read with" and "Child 'reads' out loud the story of a book to someone else (to parent, sibling, doll)." The items have been used in previous studies (Lerkkanen et al., 2006–2016; Lohvansuu et al., 2021).

References:

Lerkkanen, M.-K., Niemi, P., Poikkeus, A.-M., Poskiparta, E., Siekkinen, M., & Nurmi, J.-E. (2006–2016). *The First Steps Study* [Alkuportaat]. University of Jyväskylä.

Lohvansuu, K., Torppa, M., Ahonen, T., Eklund, K., Hämäläinen, J. A., Leppänen, P. H., & Lyytinen, H. (2021). Unveiling the mysteries of dyslexia—Lessons learned from the prospective Jyväskylä longitudinal study of dyslexia. *Brain Sciences*, *11*(4), 427. https://doi.org/10.3390/brainsci11040427

Home Numeracy Environment (HNE) and Home Literacy Environment (HLE)

The Frequency of Children's Activities in Digital Environments

There were two items on children's frequency of digital activities at home. Parents were asked to respond to the following question: "How often do the following activities happen at your home?" using a Likert scale (1 = *not at all or rarely*, 5 = *several times a day*). Items: "Child watches TV" and "Child uses mobile phone, computer, or iPad."

Reference:

Developed in the project.

HNE Informal: Games and Everyday Shared Activities

There were 17 items on informal numeracy activities adapted from LeFevre et al. (2009). Parents were asked to respond to the following question: "In the past month, how often have the following activities happened at your home?" using a Likert scale (1 = *not at all or rarely*, 5 = *several times a day*). Item examples: "Playing card games"; "Making collections"; "Playing board games with die or spinner"; "Measuring ingredients when cooking"; "Using calendars and dates."

Reference:

LeFevre, J.-A., Skwarchuk, S.-L., Smith-Chant, B. L., Fast, L., Kamawar, D., & Bisanz, J. (2009). Home numeracy experiences and children's math performance in the early school years. *Canadian Journal of Behavioural Science / Revue anadienne des sciences du comportement*, *41*(2), 55–66. https://doi.org/10.1037/a0014532

HLE Informal: Shared Reading

There were four items on shared reading activities at home. Parents were asked to respond to the following question: "How often do the following activities happen at your home?" using a Likert scale (1 = *not at all or rarely*, 5 = *several times a day*). Item examples: "Mother reads with the child"; "Father reads with the child." The items have been used in previous studies (Lerkkanen et al., 2006–2016; Lohvansuu et al., 2021).

References:

Lerkkanen, M.-K., Niemi, P., Poikkeus, A.-M., Poskiparta, E., Siekkinen, M., & Nurmi, J.-E. (2006–2016). *The First Steps Study* [Alkuportaat]. University of Jyväskylä.

Lohvansuu, K., Torppa, M., Ahonen, T., Eklund, K., Hämäläinen, J. A., Leppänen, P. H., & Lyytinen, H. (2021). Unveiling the mysteries of dyslexia—Lessons learned from the prospective Jyväskylä longitudinal study of dyslexia. *Brain Sciences*, *11*(4), 427. https://doi.org/10.3390/brainsci11040427

HNE Formal: Teaching of Numeracy Skills

There were five items on numeracy teaching activities at home adapted from Blevins-Knabe and Musun-Miller (1996). Parents were asked to respond to the following question: "In the past week, how often have the following activities happened at your home?" using a Likert scale (1 = *not at all,* 4 = *six times a week or more often*). Item examples: "You mentioned a fact such as 1 + 1 = 2 or 4 – 2 = 2 to your child"; "You tried to teach your child number order 1,2,3,4,5…"

Reference:

Blevins-Knabe, B., & Musun-Miller, L. (1996). Number use at home by children and their parents and its relationship to early mathematical performance. *Early Development and Parenting: An International Journal of Research and Practice, 5*(1), 35–45. https://doi.org/10.1002/(SICI)1099-0917(199603)5:1<35::AID-EDP113>3.0.CO;2-0

Access to Print

There were two items on access to print at home, one on library visits and another one on the amount of books at home. Regarding library visits, parents were asked to respond to the following item: "Child goes to the library with someone" using a Likert scale (1 = *not at all*, 5 = *several times a week*). Regarding the amount of books at home, parents were asked to respond to the following item: "The amount of children's books at home (own and borrowed from library)" using a Likert scale (1 = *less than 5*, 5 = *more than 100*). The items have been used in previous studies (Lerkkanen et al., 2006–2016; Lohvansuu et al., 2021).

References:

Lerkkanen, M.-K., Niemi, P., Poikkeus, A.-M., Poskiparta, E., Siekkinen, M., & Nurmi, J.-E. (2006–2016). *The First Steps Study* [Alkuportaat]. University of Jyväskylä.

Lohvansuu, K., Torppa, M., Ahonen, T., Eklund, K., Hämäläinen, J. A., Leppänen, P. H., & Lyytinen, H. (2021). Unveiling the mysteries of dyslexia—Lessons learned from the prospective Jyväskylä longitudinal study of dyslexia. *Brain Sciences, 11*(4), 427. https://doi.org/10.3390/brainsci11040427

5.5.2. PQ – Preschool Phase (T3)

BACKGROUND INFORMATION

- Who filled in the questionnaire: (1) Mother; (2) Father; (3) Someone else, who: ____
- Respondent's year of birth: ____
- Living arrangements
  - Do you live (1) Together with a spouse (married) and children; (2) Together with a spouse and children; (3) In a blended family with children; (4) As a single parent with children; (5) As some other family type, what: ____
  - Do you live in the same household with the participating child: (1) Yes; (2) No; (3) Partly, ____days/month
- Parental education:
  - Own basic education: (1) Less than comprehensive school; (2) Comprehensive school; (3) General upper secondary school

  - Own vocational education: (1) No vocational education; (2) Vocational upper secondary school; (3) Vocational college education (in Finnish "opisto"); (4) Polytechnic university; (5) University
  - Basic education of child's other guardian: (1) Less than comprehensive school; (2) Comprehensive school; (3) General upper secondary school
  - Vocational education of child's other guardian: (1) No vocational education; (2) Vocational school; (3) College education (in Finnish "opisto"); (4) Polytechnic university; (5) University
- Parental employment and occupation:
  - Employment situation: (1) Working; (2) Unemployed / partly unemployed, how long to date: ____; (3) Studying; (4) Suspended without pay (laid off), how long to date: ____; (5) Parental leave, how long to date: ____; (6) Stay-home mother/father
  - What is your occupation/profession: ____
  - What is the occupation/profession of the child's other guardian: ____
- Participating child's birthday (day, month, year): ____
- Participating child's gender: (1) Boy; (2) Girl
- Which languages does the participating child speak at home: (1) Finnish; (2) Other, which: ____
- ECEC participation: ____hours/day, ____days/week
- What is the format of ECEC: (1) Fulltime; (2) Parttime
- The ECEC history of the child: Use the table to indicate your child's ECEC history in a chronological order (use the first row as an example). Write down the child's age and tick the correct box for other characteristics.

A. Child age	B. ECEC location			C. ECEC format			D. Duration h/day			E. Duration in months
	Center based ECEC	Home based ECEC (Family day carer)	Some-where else	half a day	full day	some other format	less than 4h	5–8 h	more than 9h	duration in months
Example: 1 y 2 mo – 2 y 3 mo		x			x			x		13
1.										
2.										
3.										

Parent-Reported Learning Difficulties in Reading and Mathematics

Parents were asked if they and/or the other parent of the child had experienced learning difficulties (1) in reading or writing, and (2) in mathematics or calculation. The questionnaire included one question about their own difficulties in reading or writing, one about their own difficulties in mathematics or calculation, and the same two items concerning the other parent. The parents answered each question on a 3-point Likert scale (1 = *no difficulties*, 2 = *some difficulties*, 3 = *clear difficulties*). The items have been used in previous studies (Lerkkanen et al., 2006–2016; Lohvansuu et al., 2021).

References:

Lerkkanen, M.-K., Niemi, P., Poikkeus, A.-M., Poskiparta, E., Siekkinen, M., & Nurmi, J.-E. (2006–2016). *The First Steps Study* [Alkuportaat]. University of Jyväskylä.

Lohvansuu, K., Torppa, M., Ahonen, T., Eklund, K., Hämäläinen, J. A., Leppänen, P. H., & Lyytinen, H. (2021). Unveiling the mysteries of dyslexia—Lessons learned from the prospective Jyväskylä longitudinal study of dyslexia. *Brain Sciences, 11*(4), 427. https://doi.org/10.3390/brainsci11040427

Children's Reading and Math Motivation and Temperament

The Frequency of Children's Math Activities

There were four items of the frequency of children's math activities adapted from Blevins-Knabe & Musun-Miller (1996). Parents were asked to respond to the following question: "In the past week, how often have the following activities happened at your home?" using a Likert scale (1 = *not at all*, 4 = *six times a week or more often*). Item examples: "Child mentioned a fact such as 1 + 1 = 2 or 4 – 2 = 2"; "Child recites the number line 1,2,3,4…".

Reference:

Blevins-Knabe, B., & Musun-Miller, L. (1996). Number use at home by children and their parents and its relationship to early mathematical performance. *Early Development and Parenting: An International Journal of Research and Practice, 5*(1), 35–45. https://doi.org/10.1002/(SICI)1099-0917(199603)5:1<35::AID-EDP113>3.0.CO;2-0

The Frequency of Children's Literacy Activities

There were three items of the frequency of children's independent literacy activities. Parents were asked to respond to the following question: "How often do the following activities happen at your home?" using a Likert scale (1 = *not at all or rarely*, 5 = *several times a day*). Item examples: "Child 'reads'/ looks at / browses through books/magazines/comic books on their own"; "Child asks to be read with" and "Child 'reads' out loud the story of a book to someone else (to parent, sibling, doll)." The items have been used in previous studies (Lerkkanen et al., 2006–2016; Lohvansuu et al., 2021).

References:

Lerkkanen, M.-K., Niemi, P., Poikkeus, A.-M., Poskiparta, E., Siekkinen, M., & Nurmi, J.-E. (2006–2016). The First Steps Study [Alkuportaat]. University of Jyväskylä.

Lohvansuu, K., Torppa, M., Ahonen, T., Eklund, K., Hämäläinen, J. A., Leppänen, P. H., & Lyytinen, H. (2021). Unveiling the mysteries of dyslexia—Lessons learned from the prospective Jyväskylä longitudinal study of dyslexia. Brain Sciences, 11(4), 427. https://doi.org/10.3390/brainsci11040427

Children's Temperament

Parents reported their child's temperament with the Survey of Individual Differences of Children and Adolescents (SIDCA: Martin, 2014). SIDCA consists of 49 items that cover six temperament traits: Positive emotionality; Negative emotionality; Activity level; Distractibility; Inhibition; Sociability; Intelligence. Parents rated each item on a 7-point scale based on relative frequency of occurrence (1 = *much less than average*, 4 = *average*, 7 = *much more than average*).

Reference:

Martin, R. P. (2014). *Survey of children's individual differences* (SIDCA, Early Childhood Version). Unpublished Manual.

Home Numeracy Environment (HNE) and Home Literacy Environment (HLE)

The Frequency of Children's Activities in Digital Environments

There were two items on children's frequency of digital activities at home. Parents were asked to respond to the following question: "How often do the following activities happen at your home?" using a Likert scale (1 = *not at all or rarely,* 5 = *several times a day*). Items: "Child watches TV" and "Child uses mobile phone, computer, or iPad."

Reference:

Developed in the project.

HNE Informal: Games and Everyday Shared Activities

There were 17 items on informal numeracy activities adapted from LeFevre et al. (2009). Parents were asked to respond to the following question: "In the past month, how often have the following activities happened at your home?" using a Likert scale (1 = *not at all or rarely,* 5 = *several times a day*). Item examples: "Playing card games"; "Making collections"; "Playing board games with die or spinner"; "Measuring ingredients when cooking"; "Using calendars and dates."

Reference:

LeFevre, J.-A., Skwarchuk, S.-L., Smith-Chant, B. L., Fast, L., Kamawar, D., & Bisanz, J. (2009). Home numeracy experiences and children's math performance in the early school years. *Canadian Journal of Behavioural Science / Revue anadienne des sciences du comportement*, *41*(2), 55–66. https://doi.org/10.1037/a0014532

HLE Informal: Shared Reading

There were four items on shared reading activities at home. Parents were asked to respond to the following question: "How often do the following activities happen at your home?" using a Likert scale (1 = *not at all or rarely,* 5 = *several times a day*). Item examples: "Mother reads with the child"; "Father reads with the child." The items have been used in previous studies (Lerkkanen et al., 2006–2016; Lohvansuu et al., 2021).

References:

Lerkkanen, M.-K., Niemi, P., Poikkeus, A.-M., Poskiparta, E., Siekkinen, M., & Nurmi, J.-E. (2006–2016). *The First Steps Study* [Alkuportaat]. University of Jyväskylä.

Lohvansuu, K., Torppa, M., Ahonen, T., Eklund, K., Hämäläinen, J. A., Leppänen, P. H., & Lyytinen, H. (2021). Unveiling the mysteries of dyslexia—Lessons learned from the prospective Jyväskylä longitudinal study of dyslexia. *Brain Sciences*, *11*(4), 427. https://doi.org/10.3390/brainsci11040427

HNE Formal: Teaching of Numeracy Skills

There were five items on numeracy teaching activities at home adapted from Blevins-Knabe and Musun-Miller (1996). Parents were asked to respond to the following question: "In the past week, how often have the following activities happened at your home?" using a Likert scale (1 = *not at all,* 4 = *six times a week or more often*). Item examples: "You mentioned a fact such as 1 + 1 = 2 or 4 – 2 = 2 to your child"; "You tried to teach your child number order 1,2,3,4,5…"

Reference:

Blevins-Knabe, B., & Musun-Miller, L. (1996). Number use at home by children and their parents and its relationship to early mathematical performance. *Early Development and Parenting: An International Journal of Research and Practice, 5*(1), 35–45. https://doi.org/10.1002/(SICI)1099-0917(199603)5:1<35::AID-EDP113>3.0.CO;2-0

Access to Print

There were two items on access to print at home, one on library visits and another one on the amount of books at home. Regarding library visits, parents were asked to respond to the following item: "Child goes to the library with someone" using a Likert scale (1 = *not at all,* 5 = *several times a week*). Regarding the amount of books at home, parents were asked to respond to the following item: "The amount of children's books at home (own and borrowed from library)" using a Likert scale (1 = *less than 5,* 5 = *more than 100*). The items have been used in previous studies (Lerkkanen et al., 2006–2016; Lohvansuu et al., 2021).

References:

Lerkkanen, M.-K., Niemi, P., Poikkeus, A.-M., Poskiparta, E., Siekkinen, M., & Nurmi, J.-E. (2006–2016). *The First Steps Study* [Alkuportaat]. University of Jyväskylä.

Lohvansuu, K., Torppa, M., Ahonen, T., Eklund, K., Hämäläinen, J. A., Leppänen, P. H., & Lyytinen, H. (2021). Unveiling the mysteries of dyslexia—Lessons learned from the prospective Jyväskylä longitudinal study of dyslexia. *Brain Sciences, 11*(4), 427. https://doi.org/10.3390/brainsci11040427

Parents' Own Reading Experiences

The Frequency of Parental Leisure Reading

There were three items on parental leisure reading. Parents responded to the questions using a Likert scale (1 = *not at all,* 7 = *daily*). Items: "You read for leisure"; "Your spouse reads for leisure"; "The child sees you and your spouse reading for leisure."

Reference:

Developed in the project.

5.5.3. PQ – School Phase: Grade 1 (T5)

BACKGROUND INFORMATION

- Who filled in the questionnaire: (1) Mother; (2) Father; (3) Someone else, who: ____
- Respondent's year of birth: ____
- Are you a biological parent for the participating child: (1) Yes; (2) No
- Living arrangements
  - Do you live (1) Together with a spouse (married) and children; (2) Together with a spouse (not married) and children; (3) In a blended family with children; (4) As a single parent with children; (5) As some other family type, what: ____
  - Do you live in the same household with the participating child: (1) Yes; (2) No; (3) Partly, ____days/month
- Parental education:
  - Own basic education: (1) Less than comprehensive school; (2) Comprehensive school; (3) General upper secondary school
  - Own vocational education: (1) No vocational education; (2) Vocational school; (3) College education (in Finnish "opisto"); (4) Polytechnic university; (5) University; (6) Researcher training (Licentiate, PhD); (7) Studies ongoing, the upcoming degree: ____
  - Basic education of child's other biological parent: (1) Less than comprehensive school; (2) Comprehensive school; (3) General upper secondary school
  - Vocational education of child's other biological parent: (1) No vocational education; (2) Vocational upper secondary school; (3) Vocational college education (in Finnish "opisto"); (4) Polytechnic university; (5) University; (6) Researcher training (Licentiate, PhD); (7) Studies ongoing, the upcoming degree: ____
- Parental employment and occupation:
  - Own employment situation: (1) Working; (2) Unemployed / partly unemployed, how long to date: ____; (3) Studying; (4) Suspended without pay (laid off), how long to date: ____; (5) Parental leave, how long to date: ____; (6) Stay-home mother/father, (7) Something else, what: ____
  - Own occupation: What is your occupation/profession: ____
  - Employment situation of child's other biological parent: (1) Working; (2) Unemployed / partly unemployed, how long to date: ____; (3) Studying; (4) Suspended without pay (laid off), how long to date: ____; (5) Parental leave, how long to date: ____; (6) Stay-home mother/father, (7) Something else, what: ____
  - Occupation of child's other biological parent: What is the occupation/profession of child's other biological parent: ____
- Participating child's birthday (day, month, year): ____
- Participating child's gender: (1) Boy; (2) Girl; (3) Other
- Languages used at home:
  - What languages are spoken in your family: (1) Finnish; (2) Other, which: ____
  - What is the first language/mother tongue of the participating child's mother: (1) Finnish; (2) Other, which: ____
  - What is the first language/mother tongue of the participating child's father: (1) Finnish; (2) Other, which: ____
  - Which languages the participating child speaks at home: (1) Finnish; (2) Other, which: ____

  - What is the primary language that the participating child uses at home: (1) Finnish; (2) Other, which: ____
- ECEC and pre-primary education participation
  - How many years the child participated in ECEC before school entry: ____
  - Did the child participate in pre-primary education: (1) Yes, in ECEC center; (2) Yes, in school; (3) No
  - What was the format of pre-primary education: (1) Only pre-primary education (4 h/day); (2) Pre-primary education and additional ECEC. How many hours in addition to pre-primary education: ____h/day

Learning Difficulties in the Family

Parents were asked if they and/or the other biological parent, siblings or other close relatives of the child had experienced learning difficulties (1) in reading and/or writing, and (2) in mathematics and/or calculation. The measure consisted of eight items, including one question about their own difficulties in reading and/or writing, one about their own difficulties in mathematics and/or calculation, and the same two items concerning the other biological parent, siblings and other close relatives of the child. The parents answered each question on a 3-point Likert scale (1 = *no difficulties*, 2 = *some difficulties*, 3 = *clear difficulties*). In addition, the parents were able to choose the response "4 = *I do not know*." The scale was developed from the scale used previously in the First Steps Study (Lerkkanen et al., 2006-2016).

Reference:

Lerkkanen, M.-K., Niemi, P., Poikkeus, A.-M., Poskiparta, E., Siekkinen, M., & Nurmi, J.-E. (2006–2016). *The First Steps Study* [Alkuportaat]. University of Jyväskylä.

Children's Developmental Disabilities and Other Diagnoses

Parents were asked to report if their child had been identified as having some difficulties that affect learning using the following question "Has your child been identified as having difficulties that may affect learning? Please circle a number below and add information of who identified the difficulty: maternity clinic, doctor, someone else." The answer options were the following: 1 "No difficulties identified"; 2 "Delayed speech or language development"; 3 "Delayed motor development"; 4 "Specific language impairment"; 5 "ADHD/ADD"; 6 "Conduct disorder"; 7 "Difficulties in social situations"; 8 "Perceptual deficits"; 9 "Difficulties in emotion regulation"; 10 "Difficulties concentrating / attention focusing" and 11 "Some other difficulty, which?"

Reference:

Developed in the project.

Children's Reading and Math Motivation

The Frequency of Children's Independent Math Activities at Home

Parents were asked to evaluate how often their child spontaneously does various mathematics-related activities at home. The measure included 11 items. Two items were adapted from Blevins-Knabe and Musun-Miller (1996): "The child counts objects to figure out how many

there are" and "The child mentions number-related facts, such as '1 + 1 = 2 or 4 – 2 = 2'." Four items were adapted from Hart et al. (2016), for example: "Child draws maps or plans (e.g., for houses, castles, forts or other buildings)." The rest of the questions were adapted from the FLARE project: "The child calculates additions and subtractions"; "The child makes puzzles" and "The child builds miniatures (e.g., airplanes, dollhouses, animals)." Parents responded using a 6-point scale (1 = *not at all*, 6 = *almost daily).*

References:

Blevins-Knabe, B., & Musun-Miller, L. (1996). Number use at home by children and their parents and its relationship to early mathematical performance. *Early Development and Parenting: An International Journal of Research and Practice, 5*(1), 35–45. https://doi.org/10.1002/(SICI)1099-0917(199603)5:1<35::AID-EDP113>3.0.CO;2-0

Hart, S. A., Ganley, C. M., & Purpura, D. J. (2016). Understanding the home math environment and its role in predicting parent report of children's math skills. *PloS One, 11*(12), e0168227. https://doi.org/10.1371/journal.pone.0168227

The Frequency of Children's Independent Reading Activities at Home

Parents were asked to evaluate how often their child engages independently in reading-related activities at home. There were seven items including, for example, the following: "The child asks an adult to read to them"; "The child reads books independently" and "The child asks questions about what they have read." Items 1 to 5 were modified from items used in previous studies (Lerkkanen et al., 2006–2016; Lohvansuu et al., 2021) and items 6 and 7 were developed in the current project. Parents responded using a 4-point scale (1 = *not at all or rarely*, 4 = *once a day / on a daily basis*).

References:

Lerkkanen, M.-K., Niemi, P., Poikkeus, A.-M., Poskiparta, E., Siekkinen, M., & Nurmi, J.-E. (2006–2016). The First Steps Study [Alkuportaat]. University of Jyväskylä.

Lohvansuu, K., Torppa, M., Ahonen, T., Eklund, K., Hämäläinen, J. A., Leppänen, P. H., & Lyytinen, H. (2021). Unveiling the mysteries of dyslexia—Lessons learned from the prospective Jyväskylä longitudinal study of dyslexia. Brain Sciences, 11(4), 427. https://doi.org/10.3390/brainsci11040427

Children's Task Strategies with Homework: Task-Focused vs. Task-Avoidant Behavior

Parents were asked to report their child's task strategies (task-focused vs. task-avoidant behavior) with homework. There were 15 items altogether: five items for three domains (reading, spelling/writing, and mathematics). Item examples: "Does your child actively try to cope with difficult tasks…in reading / in writing / in mathematics?" and "Does your child easily give up trying…in reading / in writing / in mathematics?" Parents responded using a 5-point scale (1 = *not at all*, 5 = *to a high extent / very quickly*). The scale is an adaptation from the Strategy and Attribution Questionnaire (SAQ; Nurmi et al., 1995; see also Aunola et al., 1999).

References:

Aunola, K., Nurmi, J.-E., Onatsu-Arvilommi, T., & Pulkkinen, L. (1999). The role of parents' self-esteem, mastery-orientation and social background in their parenting styles. *Scandinavian Journal of Psychology, 40*(4), 307–317. https://doi.org/10.1111/1467-9450.404131

Nurmi, J.-E., Salmela-Aro, K. & Haavisto, T. (1995). The Strategy and Attribution Questionnaire: Psychometric properties. *European Journal of Psychological Assessment, 11,* 108–121.

Parent-Reported Children's Interest in Math

Parents were asked to evaluate their child's interest in math using two items (Aunola & Nurmi, 1999; Nurmi & Aunola, 2005): "My child is interested in math" and "My child likes to do math tasks." Parents responded using a 5-point scale: (1 = *not at all*, 5 = *all the time*).

References:

Aunola, K. & Nurmi, J.-E. (1999). *Task-value scale for children*. Unpublished task.

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. https://doi.org/10.1016/j.learninstruc.2005.04.009

Parent-Reported Children's Experience of Mastery in Math

Parents were asked to evaluate their child's experience of mastery regarding math using two items: "My child thinks that they are good at math" and "My child thinks that math tasks are easy." Parents responded using a 5-point scale: (1 = *not at all*, 5 = *all the time*).

Reference:

Developed in the project.

Parent-Reported Children's Self-Efficacy in Math

Parents were asked to evaluate their child's self-efficacy in math using two items: "My child believes they are able to do math tasks" and "My child believes they are able to learn new math skills." Parents responded using a 5-point scale: (1 = *not at all*, 5 = *all the time*).

Reference:

Developed in the project.

Parent-Reported Children's Math Anxiety

Parents were asked to evaluate their child's math anxiety using two items: "My child is nervous about math class" and "My child becomes anxious when they need to do math tasks." Parents responded using a 5-point scale: (1 = *not at all*, 5 = *all the time*).

Reference:

Developed in the project.

Children's Favorite Books

Parents were asked to report a few of their children's favorite books (an open-ended question).

Reference:

Developed in the project.

Home Numeracy Environment (HNE) and Home Literacy Environment (HLE)

The Frequency of Children's Activities in Digital Environments

Parents were asked to evaluate the frequency of their child's activities in digital environments. There were five items, such as "Child watches shows, series or movies (e.g., on television, streaming services, YouTube)" and "Child uses computer/mobile device to write or look for information." Parents responded using a 5-point scale (1 = *not at all or rarely*, 5 = *many times a day*).

Reference:

Developed in the project.

Children's Home Math Environment: Informal Activities

Parents were asked to evaluate how often they do math-related activities at home with their child. The measure included seven items. It was based on the measurement in Hart et al. (2016) where 48 statements were used. For example, the following statements were included: "Card games are played with the child" and "Ingredients are measured with the child while cooking." Parents responded using a 6-point scale (1 = *not at all*, 6 = *almost daily*).

Reference:

Hart, S. A., Ganley, C. M., & Purpura, D. J. (2016). Understanding the home math environment and its role in predicting parent report of children's math skills. *PloS One, 11*(12), e0168227. https://doi.org/10.1371/journal.pone.0168227

Children's Home Literacy Environment: Informal Activities

Parents were asked to evaluate how often they do reading-related activities at home with their child. The measure included six items including, for example, the following: "Mother reads a book or a magazine with the child"; "Father reads a book or a magazine with the child" and "The child is asked questions about the text that was read together." Items 1 to 4 were modified from items used in previous studies (Lerkkanen et al., 2006–2016; Lohvansuu et al., 2021) and items 5 and 6 were developed in the current project. Parents responded using a 4-point scale (1 = *not at all or rarely*, 4 = *once a day / on a daily basis*).

References:

Lerkkanen, M.-K., Niemi, P., Poikkeus, A.-M., Poskiparta, E., Siekkinen, M., & Nurmi, J.-E. (2006–2016). *The First Steps Study* [Alkuportaat]. University of Jyväskylä.

Lohvansuu, K., Torppa, M., Ahonen, T., Eklund, K., Hämäläinen, J. A., Leppänen, P. H., & Lyytinen, H. (2021). Unveiling the mysteries of dyslexia—Lessons learned from the

prospective Jyväskylä longitudinal study of dyslexia. Brain Sciences, 11(4), 427. https://doi.org/10.3390/brainsci11040427

Parental Teaching and Encouragement in Reading, Spelling, and Mathematics at Home

Parents were asked to evaluate how often they teach or encourage their child with reading, writing and mathematics. The measure included six items: "How often do you teach your child…to read / to write / math tasks?" and "How often do you encourage your child to…read independently / write independently / do math tasks independently?" Parents responded using a 5-point scale (1 = *not at all*, 5 = *daily*). The items have been adapted from Silinskas et al. (2015).

Reference:

Silinskas, G., Kiuru, N., Aunola, K., Lerkkanen, M.-K., & Nurmi, J.-E. (2015). The developmental dynamics of children's academic performance and mothers' homework-related affect and practices. *Developmental Psychology*, *51*(4), 419. https://doi.org/10.1037/a0038908

Amount of Books at Home

The parents were asked to evaluate how many children's books they have at home, including their own and borrowed books. The item was used in previous studies (Lerkkanen et al., 2006–2016; Lohvansuu et al., 2021). Parents responded using a 5-point scale (1 = *less than 5 books*, 5 = *over 100 books*).

References:

Lerkkanen, M.-K., Niemi, P., Poikkeus, A.-M., Poskiparta, E., Siekkinen, M., & Nurmi, J.-E. (2006–2016). *The First Steps Study* [Alkuportaat]. University of Jyväskylä.

Lohvansuu, K., Torppa, M., Ahonen, T., Eklund, K., Hämäläinen, J. A., Leppänen, P. H., & Lyytinen, H. (2021). Unveiling the mysteries of dyslexia—Lessons learned from the prospective Jyväskylä longitudinal study of dyslexia. Brain Sciences, 11(4), 427. https://doi.org/10.3390/brainsci11040427

Children's Homework and Parental Involvement

Children's Homework: Frequency and Amount of Time Spent

Parents were asked to report about the frequency of their child's homework using the following questions: "Does your child get homework…(a) in mathematics; (b) in language and literacy?" They responded to the items using a 5-point scale (1 = *not at all*, 5 = *daily)*. In addition, they were asked to report the amount of time (minutes per day) their child uses for homework (open-ended question).

Reference:

Developed in the project.

Parental Involvement in Children's Homework: Monitoring and Assistance

Parents were asked to report on their involvement in their child's homework with 12 items. Item examples: "How often do you make sure your child has done their homework?"; "How often do you teach your child with their homework?"; "How often do you wonder if your child has done their homework?" and "How often do you show your child you are dissatisfied if they have not done their homework?" Parents responded using a 5-point scale (1 = *never*, 5 = *always*).

Items 1, 2, 3, 4, 5, 6, 7 and 9 are from Tunkkari et al. (2021); Items 8, 10, 11 and 12 are from Gonida and Cortina (2014).

References:

Gonida, E. N., & Cortina, K. (2014). Parent involvement in homework: Relations with parent and student achievement-related motivational beliefs and achievement. *British Journal of Educational Psychology*, *84*, 376–396. https://doi.org/10.1111/bjep.12039

Tunkkari, M., Aunola, K., Hirvonen, R., Silinskas, G., & Kiuru, N. (2021). The interplay between maternal homework involvement, task-avoidance, and achievement among adolescents. *Journal of Family Psychology*, *35*(7), 863. https://doi.org/10.1037/fam0000686

Parental Homework-Related Affect

Parents were asked to evaluate how they feel when they help or advise their child with homework. Altogether four items were included: "How often do you…feel satisfied / feel helplessness / feel joy / feel frustration/stress?" Parents responded using a 5-point scale (1 = *not at all*, 5 = *to high extent / very often*). The scale is adapted to parents from Pekrun et al. (2002) and Watson et al. (1988) and was used before in the First Steps Study (Lerkkanen et al., 2006–2016).

References:

Lerkkanen, M.-K., Niemi, P., Poikkeus, A.-M., Poskiparta, E., Siekkinen, M., & Nurmi, J.-E. (2006–2016). *The First Steps Study* [Alkuportaat]. University of Jyväskylä.

Pekrun, R., Goetz, T., Titz, W., & Perry, R. P. (2002). Academic emotions in students' self-regulated learning and achievement: A program of qualitative and quantitative research. *Educational Psychologist, 37,* 91–105. https://doi.org/10.1207/S15326985EP3702_4

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS Scales. *Journal of Personality & Social Psychology*, *54*, 1063–1070. https://doi.org/10.1037/0022-3514.54.6.1063

Parental Beliefs on Child's School Achievement

Parents' beliefs regarding their children's academic performance were assessed with eight items, which parents rated on a 5-point scale (1 = *poorly*, 5 = *very well*). There were two questions, each with four achievement categories (reading, spelling/writing, mathematics, and school in general). The questions were as follows: "How well does your child succeed in school…in reading / in writing / in mathematics / at school in general?" and "How well do you think your child will succeed later in school…in reading / in writing / in mathematics / at school

in general?" The scale was developed based on Nicholls (1978) and the scales used in the First Steps Study (Lerkkanen et al., 2006–2016)

References:

Lerkkanen, M.-K., Niemi, P., Poikkeus, A.-M., Poskiparta, E., Siekkinen, M., & Nurmi, J.-E. (2006–2016). *The First Steps Study* [Alkuportaat]. University of Jyväskylä.

Nicholls, J. G. (1978). The development of the concepts of effort and ability, perception of academic attainment and the understanding that difficult tasks require more ability. *Child Development, 49,* 800–814. https://doi.org/10.2307/1128250

Parents' Own Reading and Math Skills and Experiences

Adult Reading History Questionnaire

Parents' reading history was measured with the Adult Reading History Questionnaire (ARHQ; Lefly & Pennington, 2000). The measure includes 26 questions, of which 23 were used in this research. The sequence of the questions was changed to support responding by sorting the items into three categories: (1) Parent's childhood reading skills and difficulties, (e.g., "How much difficulty did you have learning to read in elementary school?"), (2) Parent's current reading habits (e.g., "How much reading do you do for pleasure?") and (3) Parent's current reading skills (e.g., "How would you compare your current reading speed to that of others of the same age and education?"). The scales in ARHQ vary from item to item, but lower scores indicate less difficulty.

Reference:

Lefly, D. L., & Pennington, B. F. (2000). Reliability and validity of the adult reading history questionnaire. *Journal of Learning Disabilities*, *33*(3), 286–296. https://doi.org/10.1177/002221940003300306

Parents' Childhood Math Learning Difficulties

Parents were asked to evaluate if they had had difficulties to learn different math content. There were five items, such as "Did you have difficulties in learning multiplications?" and "Did you have difficulties in learning units and measuring (e.g., mm, cm, dm…)?" The items were developed in the project based on the work by Nguyen et al. (2022), Stefansson et al. (2014), and Sury and Gaab (2020). Parents were asked to respond using a 5-point scale (1 = *not at all*, 5 = *to a high extent).*

References:

Nguyen, T. Q., Martinez-Lincoln, A., & Cutting, L. E. (2022). Tracking familial history of reading and math difficulties in children's academic outcomes. *Frontiers in Psychology*, *12,* 710380. https://doi.org/10.3389/fpsyg.2021.710380

Stefansson, H., Meyer-Lindenberg, A., Steinberg, S., Magnusdottir, B., Morgen, K., Arnarsdottir, S., Bjornsdottir, G., Bragi Walters, G., Jonsdottir, G. A., Doyle, O. M., Tost, H., Grimm, O., Kristjansdottir, S., Snorrason, H., Davidsdottir, S. R., Gudmundsson, L. J., Jonsson, G. F, Stefansdottir, B., Helgadottir, I., Haraldsson, M., Jonsdottir, B.,... & Tost, H. (2014). CNVs conferring risk of autism or schizophrenia

affect cognition in controls. *Nature*, *505*(7483), 361. https://doi.org/10.1038/nature12818

Sury, D., & Gaab, N. (2020). The Adult Arithmetic History Questionnaire. https://psyarxiv.com/zt6ku/

Parents' Current Math Difficulties in Everyday Situations

Parents were asked to report if they experience difficulties in everyday math. There were 15 statements, and the items were developed in the project (Koponen, 2020), but similar items can be found, for example, in Sury and Gaab (2020). The statements included the following items: "It's difficult for me to estimate that how much the purchases approximately cost"; "It's difficult for me to understand bus/train timetables" and "It's difficult for me to calculate the needed amount of wallpaper rolls if I were to wallpaper the bedroom." Parents were asked to respond using a 5-point scale (1 = *applies very poorly*, 5 = *applies very well*).

Reference:

Koponen, 2020. Unpublished questionnaire.

Sury, D., & Gaab, N. (2020). The Adult Arithmetic History Questionnaire. https://psyarxiv.com/zt6ku/

Parents' Need for Extra Support for Learning Math During their Early School Years

Parents were asked to evaluate how much extra help they needed to learn math during the early school years. There was one item and parents were asked to respond using a 5-point scale (1 = *not at all*, 5 = *to a high extent)*.

Reference:

Developed in the project.

Parents' Self-Concept of Ability in Reading, Spelling and Math

Parents were asked to evaluate their own reading, spelling and math skills using six items, two for each skill. The items are based on Nicholls (1978) and adapted from the ones used in the First Steps Study (Lerkkanen et al., 2006–2016). In the first set of three items the parents were asked to evaluate their skills compared to those of others of the same age. The first three items were as follows: "How good skills do you think you have in…reading/spelling/mathematics?" Parents were asked to respond using a 5-point scale (1 = *below average*, 5 = *above average*). The second set of three items were as follows: "Please remember back to your school years. How good were you then in comparison to your classmates at… reading/spelling/mathematics?" Parents were asked to respond using a 5-point scale (1 = *poor / not very good*, 5 = *very good*).

References:

Lerkkanen, M.-K., Niemi, P., Poikkeus, A.-M., Poskiparta, E., Siekkinen, M., & Nurmi, J.-E. (2006–2016). *The First Steps Study* [Alkuportaat]. University of Jyväskylä.

Nicholls, J. G. (1978). The development of the concepts of effort and ability, perception of academic attainment and the understanding that difficult tasks require more ability. *Child Development, 49*, 800–814. https://doi.org/10.2307/1128250

5.5.4. PQ – School Phase: Grade 2 (T6)

BACKGROUND INFORMATION

- Who filled in the questionnaire: (1) Mother; (2) Father; (3) Someone else, who: ____
- Respondent's year of birth: ____
- Are you a biological parent for the participating child: (1) Yes; (2) No
- Living arrangements
  - Do you live (1) Together with a spouse (married or not married) and children; (2) In a blended family with children; (3) As a single parent with children; (4) As some other family type, what: ____
  - Do you live in the same household with the participating child: (1) Yes; (2) No; (3) Partly, ____days/month
- Parental education:
  - Own basic education: (1) Less than comprehensive school; (2) Comprehensive school; (3) General upper secondary school
  - Own vocational education: (1) No vocational education; (2) Vocational school; (3) College education (in Finnish "opisto"); (4) Polytechnic university, lower degree; (5) Polytechnic university, higher degree; (6) University, bachelor; (7) University, masters; (8) Researcher training (Licentiate, PhD); (9) Some other, which: ____; (10) Studies ongoing, the upcoming degree: ____
  - Basic education of child's other biological parent: (1) Less than comprehensive school; (2) Comprehensive school; (3) General upper secondary school
  - Vocational education of child's other biological parent: (1) No vocational education; (2) Vocational upper secondary school; (3) Vocational college education (in Finnish "opisto"); (4) Polytechnic university, lower degree; (5) Polytechnic university, higher degree; (6) University, bachelor; (7) University, masters; (8) Researcher training (Licentiate, PhD); (9) Some other, which: ____; (10) Studies ongoing, the upcoming degree: ____
- Parental employment and occupation:
  - Own employment situation: (1) Working; (2) Entrepreneur; (3) Unemployed; (4) Studying; (5) Suspended without pay (laid off); (6) Parental leave; (7) Stay-home mother/father; (8) Something else, what: ____
  - Own occupation: What is your occupation/profession: ____
  - Employment situation of the child's other biological parent: (1) Working; (2) Entrepreneur; (3) Unemployed; (4) Studying; (5) Suspended without pay (laid off); (6) Parental leave; (7) Stay-home mother/father; (8) Something else, what: ____
  - Occupation of the child's other biological parent: What is the occupation/profession of child's other biological parent: ____
- Participating child's birthday (day, month, year): ____
- Participating child's gender: (1) Boy; (2) Girl; (3) Other
- Languages used at home:
  - What languages are spoken in your family: (1) Finnish; (2) Other, which: ____
  - What is the first language/mother tongue of the participating child's mother: (1) Finnish; (2) Other, which: ____
  - What is the first language/mother tongue of the participating child's father: (1) Finnish; (2) Other, which: ____
  - Which languages the participating child speaks at home: (1) Finnish; (2) Other, which: ____

- What is the primary language that the participating child uses at home: (1) Finnish; (2) Other, which: ____
- ECEC and pre-primary education participation
  - How many years the child participated in ECEC before school entry: ____
  - Did the child participate in pre-primary education: (1) Yes, in ECEC center; (2) Yes, in school; (3) No
  - What was the format of pre-primary education: (1) Only pre-primary education (4 h/day); (2) Pre-primary education and additional ECEC. How many hours in addition to pre-primary education: ____h/day

Effects of COVID-19 Pandemic on Children's Schooling

Absenteeism: The parents were asked to evaluate how many days their child has been absent during Grade 2 (open question).

Receiving support: The parents were asked to evaluate if their child received enough support for schoolwork during the period they were absent on the following scale: (1) Yes; (2) Somewhat but not enough; (3) No. In addition, an open-ended question was included: "If you responded 2 or 3, could you please explain why: ____."

The impacts of COVID-19 pandemic on children's skills and well-being: The parents were asked to evaluate if the pandemic affected their child's development in (a) reading, spelling and writing skills; (b) math skills; (c) foreign language learning; (d) social skills; I friendships; (f) school motivation; (g) child's view of I as learner; (h) well-being. Parents responded using a 5-point scale: (1 = *negatively*, 5 = *positively*). In addition, an open-ended question was included: "Are there other things that were affected by COVID-19 in your child's life: ____."

Reference:

Developed in the project.

Children's Reading and Math Motivation

The Frequency of Children's Independent Math Activities at Home

Parents were asked to evaluate how often their child spontaneously does various mathematics-related activities at home. The measure included 10 items. One item was adapted from Blevins-Knabe and Musun-Miller (1996): "The child mentions number-related facts, such as '1 + 1 = 2 or 4 – 2 = 2'." Four items were adapted from Hart et al. (2016), for example: "Child draws maps or plans (e.g., for houses, castles, forts or other buildings)." The rest of the questions were adapted from the FLARE project: "The child calculates additions and subtractions"; "The child makes puzzles" and "The child builds miniatures (e.g., airplanes, dollhouses, animals)." Parents responded using a 6-point scale (1 = *not at all,* 6 = *almost daily).*

References:

Blevins-Knabe, B., & Musun-Miller, L. (1996). Number use at home by children and their parents and its relationship to early mathematical performance. *Early Development and Parenting: An International Journal of Research and Practice, 5*(1), 35-45. https://doi.org/10.1002/(SICI)1099-0917(199603)5:1<35::AID-EDP113>3.0.CO;2-0

Hart, S. A., Ganley, C. M., & Purpura, D. J. (2016). Understanding the home math environment and its role in predicting parent report of children's math skills. *PloS One, 11*(12), e0168227. https://doi.org/10.1371/journal.pone.0168227

The Frequency of Children's Independent Reading Activities at Home

Parents were asked to evaluate how often their child engages independently in reading-related activities at home. There were eight items including, for example, the following: "The child asks an adult to read to them"; "The child reads books independently (traditional or e-books)" and "The child asks questions about what they have read." Items 1 to 5 were modified from items used in previous studies (Lerkkanen et al., 2006–2016; Lohvansuu et al., 2021) and items 6 to 8 were developed in the current project. Parents responded using a 4-point scale (*1 = not at all or rarely, 4 = once a day / on a daily basis*).

References:

Lerkkanen, M.-K., Niemi, P., Poikkeus, A.-M., Poskiparta, E., Siekkinen, M., & Nurmi, J.-E. (2006–2016). The First Steps Study [Alkuportaat]. University of Jyväskylä.

Lohvansuu, K., Torppa, M., Ahonen, T., Eklund, K., Hämäläinen, J. A., Leppänen, P. H., & Lyytinen, H. (2021). Unveiling the mysteries of dyslexia—Lessons learned from the prospective Jyväskylä longitudinal study of dyslexia. Brain Sciences, 11(4), 427. https://doi.org/10.3390/brainsci11040427

Children's Task Strategies with Homework: Task-Focused vs. Task-Avoidant Behavior

Parents were asked to report their child's task strategies (task-focused vs. task-avoidant behavior) with homework. There were nine items altogether: three items for three domains (reading, spelling/writing, and mathematics). The items were: "If difficulties emerge in the task, does the child easily start to do something else…in reading / in writing / in mathematics?"; "Does your child actively try to cope with difficult tasks…in reading / in writing / in mathematics?" and "Does your child easily give up trying…in reading / in writing / in mathematics?" Parents responded using a 5-point scale (1 = *not at all*, 5 = *to a high extent / very quickly*). The scale is an adaptation from the Strategy and Attribution Questionnaire (SAQ; Nurmi et al., 1995; see also Aunola et al., 1999).

References:

Aunola, K., Nurmi, J.-E., Onatsu-Arvilommi, T., & Pulkkinen, L. (1999). The role of parents' self-esteem, mastery-orientation and social background in their parenting styles. *Scandinavian Journal of Psychology, 40*(4), 307–317. https://doi.org/10.1111/1467-9450.404131

Nurmi, J.-E., Salmela-Aro, K. & Haavisto, T. (1995). The Strategy and Attribution Questionnaire: Psychometric properties. *European Journal of Psychological Assessment, 11,* 108–121.

Parent-Reported Children's Interest in Reading and Writing

Parents were asked to evaluate their child's interest in reading and writing using four items (Aunola & Nurmi, 1999; Nurmi & Aunola, 2005): two for reading and two for writing. The

questions were as follows: “My child is interested in… reading / writing” and “My child likes to do…reading-related homework / writing-related homework.” Parents responded using a 5-point scale (1 = *not at all*, 5 = *all the time)*.

References:

Aunola, K. & Nurmi, J.-E. (1999). *Task-value scale for children*. Unpublished task.

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.* https://doi.org/10.1016/j.learninstruc.2005.04.009

Parent-Reported Children’s Interest in Math

Parents were asked to evaluate their child’s interest in math using two items (Aunola & Nurmi, 1999; Nurmi & Aunola, 2005): “My child is interested in math” and “My child likes to do math tasks.” Parents responded using a 5-point scale: (1 = *not at all*, 5 = *all the time*).

References:

Aunola, K. & Nurmi, J.-E. (1999). *Task-value scale for children*. Unpublished task.

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.* https://doi.org/10.1016/j.learninstruc.2005.04.009

Parent-Reported Children’s Experience of Mastery in Math

Parents were asked to evaluate their child’s experience of mastery regarding math using two items: “My child thinks that they are good at math” and “My child thinks that math tasks are easy.” Parents responded using a 5-point scale: (1 = *not at all*, 5 = *all the time*).

Reference:

Developed in the project.

Parent-Reported Children’s Self-Efficacy in Math

Parents were asked to evaluate their child’s self-efficacy in math using two items: “My child believes they are able to do math tasks” and “My child believes they are able to learn new math skills.” Parents responded using a 5-point scale: (1 = *not at all*, 5 = *all the time*).

Reference:

Developed in the project.

Parent-Reported Children’s Math Anxiety

Parents were asked to rate whether their child feels nervous or anxious about mathematics. The measure consisted of 11 statements adapted from Francis et al. (2020). Item examples: “My child worries of not being good at mathematics”; “My child sleeps badly because they are worried about mathematics” and “My child worries about making mistakes in mathematics.” Parents responded using a 4-point scale *(*1 = *not at all*, 4 = *all the time)*.

Reference:

Francis, D., Nation, K., & McArthur, G. (2020). *The Macquarie Oxford Reading Anxiety Test-Parents (MoRAT-P).* Motif.

Parent-Reported Children's Reading Anxiety

Parents were asked to rate whether their child feels nervous or anxious about reading. The measure consisted of 11 statements adapted from Francis et al. (2020). Item examples: "My child worries of not being good at reading"; "My child sleeps badly because they are worried about reading" and "My child worries about making mistakes in reading." Parents responded using a 4-point scale (1 = *not at all*, 4 = *all the time*).

Reference:

Francis, D., Nation, K., & McArthur, G. (2020). *The Macquarie Oxford Reading Anxiety Test-Parents (MoRAT-P).* Motif.

Home Numeracy Environment (HNE) and Home Literacy Environment (HLE)

The Frequency of Children's Activities in Digital Environments

Parents were asked to evaluate the frequency of their child's activities in digital environments. There were five items, such as "Child watches shows, series or movies (e.g., on television, streaming services, YouTube)" and "Child uses a computer/mobile device to write or look for information." Parents responded using a 5-point scale (1 = *not at all or rarely*, 5 = *many times a day*).

Reference:

Developed in the project.

Children's Home Math Environment: Informal Activities

Parents were asked to evaluate how often they do math-related activities at home with their child. The measure included seven items. It was based on the measurement in Hart et al. (2016) where 48 statements were used. For example, the following statements were included: "Card games are played with the child" and "Ingredients are measured with the child while cooking." Parents responded using a 6-point scale (1 = *not at all*, 6 = *almost daily).*

Reference:

Hart, S. A., Ganley, C. M., & Purpura, D. J. (2016). Understanding the home math environment and its role in predicting parent report of children's math skills. *PloS One, 11*(12), e0168227. https://doi.org/10.1371/journal.pone.0168227

Children's Home Literacy Environment: Informal Activities

Parents were asked to evaluate how often they do reading-related activities at home with their child. The measure included six items, including, for example, the following: "Mother reads a book or a magazine with the child (traditional or e-magazines/-books)"; "Father reads a book

or a magazine with the child (traditional or e-magazines/-books)" and "The child is asked questions about the text that was read together." Items 1 to 4 were modified from items used in previous studies (Lerkkanen et al., 2006–2016; Lohvansuu et al., 2021) and items 5 and 6 were developed in the current project. Parents responded using a 4-point scale (1 = *not at all or rarely*, 4 = *once a day / on a daily basis*).

References:

Lerkkanen, M.-K., Niemi, P., Poikkeus, A.-M., Poskiparta, E., Siekkinen, M., & Nurmi, J.-E. (2006–2016). The First Steps Study [Alkuportaat]. University of Jyväskylä.

Lohvansuu, K., Torppa, M., Ahonen, T., Eklund, K., Hämäläinen, J. A., Leppänen, P. H., & Lyytinen, H. (2021). Unveiling the mysteries of dyslexia—Lessons learned from the prospective Jyväskylä longitudinal study of dyslexia. Brain Sciences, 11(4), 427. https://doi.org/10.3390/brainsci11040427

Parental Teaching and Encouragement in Reading, Spelling, and Mathematics at Home

Parents were asked to evaluate how often they teach or encourage their child with reading, writing and mathematics. The measure included six items: "How often do you teach your child… to read / to write / math tasks?" and "How often do you encourage your child to…read independently / write independently / do math tasks independently?" Parents responded using a 5-point scale (1 = *not at all*, 5 = *daily*). The items have been adapted from Silinskas et al. (2015).

Reference:

Silinskas, G., Kiuru, N., Aunola, K., Lerkkanen, M.-K., & Nurmi, J.-E. (2015). The developmental dynamics of children's academic performance and mothers' homework-related affect and practices. *Developmental Psychology*, *51*(4), 419. https://doi.org/10.1037/a0038908

Children's Homework and Parental Involvement

Children's Homework: Frequency and Amount of Time Spent

Parents were asked to report about the frequency of their child's homework using the following questions: "Does your child get homework…(a) in mathematics; (b) in language and literacy?" They responded to the items using a 5-point scale (1 = *not at all*, 5 = *daily).* In addition, they were asked to report the amount of time (minutes per day) their child uses for homework (open-ended question).

Reference:

Developed in the project.

Children's Homework: Amount of Time the Parents Help with Homework

Parents were asked to report the amount of time (minutes per day) they use to help their child with homework (open-ended question).

Reference:

Developed in the project.

Parental Involvement in Children's Homework: Monitoring and Assistance

Parents were asked to report on their involvement in their child's homework with five items. Item examples: "How often do you make sure your child has done their homework?" and "How often do you help or guide your child with mathematics homework?" Parents responded using a 5-point scale (1 = *never*, 5 = *always*). The items have been used in previous study (Tunkkari et al., 2021).

Reference:

Tunkkari, M., Aunola, K., Hirvonen, R., Silinskas, G., & Kiuru, N. (2021). The interplay between maternal homework involvement, task-avoidance, and achievement among adolescents. *Journal of Family Psychology*, *35*(7), 863. https://doi.org/10.1037/fam0000686

Parental Homework-Related Affect

Parents were asked to evaluate how they feel when they help or advise their child with homework. Altogether four items were included: "How often do you…feel satisfied / feel helplessness / feel joy / feel frustration/stress?" Parents responded using a 5-point scale (1 = *not at all*, 5 = *to high extent / very often*). The scale is adapted to parents from Pekrun et al. (2002) and Watson et al. (1988) and was used before in the First Steps Study (Lerkkanen et al., 2006–2016).

References:

Lerkkanen, M.-K., Niemi, P., Poikkeus, A.-M., Poskiparta, E., Siekkinen, M., & Nurmi, J.-E. (2006–2016). *The First Steps Study* [Alkuportaat]. University of Jyväskylä.

Pekrun, R., Goetz, T., Titz, W., & Perry, R. P. (2002). Academic emotions in students' self-regulated learning and achievement: A program of qualitative and quantitative research. *Educational Psychologist, 37*, 91–105. https://doi.org/10.1207/S15326985EP3702_4

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS Scales. *Journal of Personality & Social Psychology, 54*, 1063–1070. https://doi.org/10.1037/0022-3514.54.6.1063

Parental Reasons to Support Child in Homework

Parents were asked to evaluate why they are involved in their child's homework with 11 statements. Item examples: "I'm involved in my child's homework…because it might be useful for them in the future" / "…to make learning nicer for them" / "…because I would feel dissatisfied with myself otherwise" and "…because they need help." In addition, the parents could add other reasons that were not included in the statements (open-ended question). Parents responded using a 4-point scale (1 = *applies very poorly*, 4 = *applies very well*). The items were developed for parents on the basis of mastery orientation scales (see e.g., Falanga et al., 2023; Midgley et al., 2000).

References:

Falanga, K., Gonida, E., & Stamovlasis, D. (2023). Predicting different types of parental involvement in children's homework: the role of parent motivational beliefs and parent affect. *European Journal of Psychology of Education, 38*(1), 249–268. https://doi.org/10.1007/s10212-022-00613-0

Midgley, C., Maehr, M. L., Hruda, L., Anderman, E. M., Anderman, L., Freeman, K. E., Gheen, M., Kaplan, A., Kumar, R., Middleton, M. J., Nelson, J., Roeser, R., & Urdan, T. (2000). *Manual for the Patterns of Adaptive Learning Scales (PALS).* University of Michigan.

Parents' Own Reading and Math Skills and Experiences

Adult Math History Questionnaire – Short Version

Parents' math history was assessed with six questions that were adapted from three different measures of adult math/arithmetic history questionnaires that were originally based on the classic Adult Reading History Questionnaire (Lefly & Pennington, 2000). Out of the six items used in this research, five are shared with the AMHQ (Nguyen et al., 2022), AAHQ (Sury & Gaab, 2020), and Stefansson et al.'s (2006) scales. The last question, "Did you experience any difficulties learning the multiplication table in elementary school?", was included in Stefansson et al.'s (2006) measure. The items were as follows: "Did you experience any difficulties in learning math in elementary school?"; "How do you rate your math skills now compared to people your age with a comparable education level?"; "What is your current attitude towards math?"; "If you compare your math skills to people of your education level and age, how would you situate yourself?" and "How much help did you need with math in primary school?" The items were rated on a 5-point scale that varied from item to item, but lower scores indicate less difficulty.

References:

Lefly, D. L., & Pennington, B. F. (2000). Reliability and validity of the adult reading history questionnaire. *Journal of Learning Disabilities*, *33*(3), 286–296. https://doi.org/10.1177/002221940003300306

Nguyen, T. Q., Martinez-Lincoln, A., & Cutting, L. E. (2022). Tracking familial history of reading and math difficulties in children's academic outcomes. *Frontiers in Psychology*, *12*, 710380. https://doi.org/10.3389/fpsyg.2021.710380

Stefansson, H., Meyer-Lindenberg, A., Steinberg, S., Magnusdottir, B., Morgen, K., Arnarsdottir, S., Bjornsdottir, G., Bragi Walters, G., Jonsdottir, G. A., Doyle, O. M., Tost, H., Grimm, O., Kristjansdottir, S., Snorrason, H., Davidsdottir, S. R., Gudmundsson, L. J., Jonsson, G. F, Stefansdottir, B., Helgadottir, I., Haraldsson, M., Jonsdottir, B.,... & Tost, H. (2014). CNVs conferring risk of autism or schizophrenia affect cognition in controls. *Nature*, *505*(7483), 361. https://doi.org/10.1038/nature12818

Sury, D., & Gaab, N. (2020). The Adult Arithmetic History Questionnaire. https://psyarxiv.com/zt6ku/

Parents' Beliefs and Confidence about Math

Parents were asked to evaluate their own math beliefs and confidence by responding to six items. The items are based on Beliefs and Confidence Survey (Chen et al., 2014) (i.e., "I'm not a 'math person'" and "I like coming up with creative ways to solve math problems"). Four items were derived from the Beliefs About Mathematics and Teaching survey (MacGyvers et al., 1993) (e.g., "When my answer to a math problem doesn't match someone else's, I usually assume that my answer is wrong" and "I enjoy encountering situations in my everyday life [e.g., sewing, carpentry, finances] that require me to use math to solve problems"). Parents responded using a 5-point scale (1 = *applies very poorly*, 5 = *applies very well*).

References:

Chen, J. Q., McCray, J., Adams, M., & Leow, C. (2014). A survey study of early childhood 1teachers' beliefs and confidence about teaching early math. *Early Childhood Education Journal*, *42*, 367–377. https://doi.org/10.1007/s10643-013-0619-0

MacGyvers, Stipek, Salmon & Bogard (1993). *The Beliefs About Mathematics and Teaching*. Unpublished research survey. UCLA Grad. School of Education.

Parents' Math Anxiety

Parents were asked to evaluate their own math anxiety with questions that were adapted from the measure created by Edwards et al. (2021). Originally the questions were related to reading, but they were adapted for math in this research. In addition, there were originally 10 items, of which nine were used in this research, and the questions were turned into statements. Item examples: "I fear people will correct my mistakes when I do math tasks"; "I avoid doing math tasks" and "I have math anxiety." Parents responded using a 5-point scale (1 = *applies very well*, 5 = *applies very poorly*).

Reference:

Edwards, A., Daucourt, M. C., Hart, S., & Schatschneider, C. (2021). Measuring Reading Anxiety in College Students. *Reading and Writing 36*, 1145–1180. https://doi.org/10.1007/s11145-022-10324-z

Parents' Reading Anxiety

Parents were asked to evaluate their own reading anxiety with a measure that was created by Edwards et al. (2021). There were originally 10 items, of which nine were used in this study, and the questions were turned into statements. Item examples: "I feel uncomfortable reading out loud in front of people"; "I worry I don't understand what I read" and "Reading a book or textbook feels like an insurmountable task." One question from the original measure was excluded from this research: "Does looking through/reading a textbook for class feel overwhelming?" Parents responded using a 5-point Likert scale (1 = *applies very well*, 5 = *applies very poorly*).

Reference:

Edwards A. A., Daucourt M. C., Hart S. A., Schatschneider C. (2022). Measuring reading anxiety in college students. *Reading and Writing*, *36*(5), 1145–1180. https://doi.org/10.1007/s11145-022-10324-z

5.5.5. PQ – School Phase: Grade 3 (T7)

BACKGROUND INFORMATION

- Who filled in the questionnaire: (1) Mother; (2) Father; (3) Someone else, who: ____
- Respondent's year of birth: ____
- Living arrangements
  - Do you live (1) Together with a spouse (married or not married) and children; (2) In a blended family with children; (3) As a single parent with children; (4) As some other family type, what: ____
  - Do you live in the same household with the participating child: (1) Yes; (2) No; (3) Partly, ____days/month
- Participating child's birthday: ____
- Children of the family:
  - How many children altogether live in your household at the moment? ____
  - The birth years of the children of the family: ____

Learning Difficulties in the Family

Parents were asked if they and/or the other biological parent, siblings or other close relatives of the child had experienced learning difficulties (1) in reading and/or writing, and (2) in mathematics and/or calculation. The measure consisted of eight items, including one question about their own difficulties in reading and/or writing, one about their own difficulties in mathematics and/or calculation, and the same two items concerning the other biological parent, siblings and other close relatives of the child. The parents answered each question on a 3-point Likert scale (1 = *no difficulties*, 2 = *some difficulties*, 3 = *clear difficulties*). In addition, the parents were able to choose the response "4 = *I do not know*." The scale was developed from the scale used previously in the First Steps Study (Lerkkanen et al., 2006–2016).

Reference:

Lerkkanen, M.-K., Niemi, P., Poikkeus, A.-M., Poskiparta, E., Siekkinen, M., & Nurmi, J.-E. (2006–2016). *The First Steps Study* [Alkuportaat]. University of Jyväskylä.

Children's Developmental Disabilities and Other Diagnoses

Parents were asked to report if their child had been identified as having some difficulties that affect learning using the following question: "Has your child been identified as having difficulties that may affect learning? Please circle a number below and add information of who identified the difficulty: maternity clinic, doctor, someone else." The answer options were the following: 1 "No difficulties identified"; 2 "Delayed speech or language development"; 3 "Delayed motor development"; 4 "Specific language impairment"; 5 "ADHD/ADD"; 6 "Conduct disorder"; 7 "Difficulties in social situations"; 8 "Perceptual deficits"; 9 "Difficulties in emotion regulation"; 10 "Difficulties concentrating / attention focusing" and 11 "Some other difficulty, which?"

Reference:

Developed in the project.

Children's Sleep: Amount and Sufficiency

Parents were asked to evaluate children's sleep with three questions: To the first item, "How many hours does your child generally sleep in a night?", the parents responded using a 6-point scale (1 = *more than 11 hours*, 6 = *less than 5 hours*). To the two other questions: "Do you think your child gets enough sleep?" and "Does your child feel tired during the day?", the parents responded using a 5-point scale (1 = *never*, 5 = *all the time* [6–7 times/week]).

Developed in the project based on the literature of some of the key assessment categories of child sleep (See, e.g., Sen & Spruyt, 2020).

Reference:

Sen, T., & Spruyt, K. (2020). Pediatric sleep tools: An updated literature review. *Frontiers in Psychiatry, 11*, 317. https://doi.org/10.3389/fpsyt.2020.00317

In addition, the following Background Information was included in the parental questionnaire for those parents who had not filled in the parental questionnaire in Grade 1 or Grade 2:

- Are you a biological parent for the participating child: (1) Yes; (2) No
- Parental education:
  - Own basic education: (1) Less than comprehensive school; (2) Comprehensive school; (3) General upper secondary school
  - Own vocational education: (1) No vocational education; (2) Vocational upper secondary school; (3) Vocational college education (in Finnish "opisto"); (4) Polytechnic university, lower degree; (5) Polytechnic university, higher degree; (6) University, bachelor; (7) University, masters; (8) Researcher training (Licentiate, PhD); (9) Some other, which: ____; (10) Studies ongoing, the upcoming degree: ____
  - Basic education of child's other biological parent: (1) Less than comprehensive school; (2) Comprehensive school; (3) General upper secondary school
  - Vocational education of the child's other biological parent: (1) No vocational education; (2) Vocational school; (3) College education (in Finnish "opisto"); (4) Polytechnic university, lower degree; (5) Polytechnic university, higher degree; (6) University, bachelor; (7) University, masters; (8) Researcher training (Licentiate, PhD); (9) Some other, which: ____; (10) Studies ongoing, the upcoming degree: ____
- Parental employment and occupation:
  - Own employment situation: (1) Working; (2) Entrepreneur; (3) Unemployed; (4) Studying; (5) Suspended without pay (laid off); (6) Parental leave; (7) Stay-home mother/father; (8) Something else, what: ____
  - Own occupation: What is your occupation/profession: ____
  - Employment situation of the child's other biological parent: (1) Working; (2) Entrepreneur; (3) Unemployed; (4) Studying; (5) Suspended without pay (laid off); (6) Parental leave; (7) Stay-home mother/father; (8) Something else, what: ____
  - Occupation of the child's other biological parent: What is the occupation/profession of child's other biological parent: ____
- Participating child's gender: (1) Boy; (2) Girl; (3) Other
- Languages used at home:
  - What languages are spoken in your family: (1) Finnish; (2) Other, which: ____
  - What is your first language/mother tongue: (1) Finnish; (2) Other, which: ____

- Which languages the participating child speaks at home: (1) Finnish; (2) Other, which: ____
- What is the primary language that the participating child uses at home: (1) Finnish; (2) Other, which: ____

Children's Reading and Math Motivation

The Frequency of Children's Independent Math Activities at Home

Parents were asked to evaluate how often their child spontaneously does various mathematics related activities at home. The measure included 10 items. One item was adapted from Blevins-Knabe and Musun-Miller (1996): "The child mentions number related facts, such as '1 + 1 = 2 or 4 – 2 = 2'." Four items were adapted from Hart et al. (2016), for example: "Child draws maps or plans (e.g., for houses, castles, forts or other buildings)." The rest of the questions were adapted from the FLARE project: "The child calculates additions and subtractions"; "The child makes puzzles" and "The child builds miniatures (e.g., airplanes, dollhouses, animals)." Parents responded using a 6-point scale (1 = *not at all,* 6 = *almost daily).*

References:

Blevins-Knabe, B., & Musun-Miller, L. (1996). Number use at home by children and their parents and its relationship to early mathematical performance. *Early Development and Parenting: An International Journal of Research and Practice, 5*(1), 35–45. https://doi.org/10.1002/(SICI)1099-0917(199603)5:1<35::AID-EDP113>3.0.CO;2-0

Hart, S. A., Ganley, C. M., & Purpura, D. J. (2016). Understanding the home math environment and its role in predicting parent report of children's math skills. *PloS One, 11*(12), e0168227. https://doi.org/10.1371/journal.pone.0168227

Children's Task Strategies with Homework: Task-Focused vs. Task-Avoidant Behavior

Parents were asked to report their child's task strategies (task-focused vs. task-avoidant behavior) with homework. There were six items altogether, with three items for two domains, (1) language and literacy and (2) mathematics. Items: "If difficulties emerge in the task, does the child easily start to do something else…in language and literacy / in mathematics?"; "Does your child actively try to cope with difficult tasks…in language and literacy / in mathematics?" and "Does your child easily give up trying… in language and literacy / in mathematics?" Parents responded using a 5-point scale (1 = *not at all,* 5 = *to a high extent / very quickly*). The scale is an adaptation from the Strategy and Attribution Questionnaire (SAQ; Nurmi et al., 1995; see also Aunola et al., 1999).

References:

Aunola, K., Nurmi, J.-E., Onatsu-Arvilommi, T., & Pulkkinen, L. (1999). The role of parents' self-esteem, mastery-orientation and social background in their parenting styles. *Scandinavian Journal of Psychology, 40*(4), 307–317. https://doi.org/10.1111/1467-9450.404131

Nurmi, J.-E., Salmela-Aro, K. & Haavisto, T. (1995). The Strategy and Attribution Questionnaire: Psychometric properties. *European Journal of Psychological Assessment, 11,* 108–121.

Parent-Reported Children's Interest in Math

Parents were asked to evaluate their child's interest in math using two items (Aunola & Nurmi, 1999; Nurmi & Aunola, 2005): "My child is interested in math" and "My child likes to do math tasks." Parents responded using a 5-point scale: (1 = *not at all*, 5 = *all the time*).

References:

Aunola, K. & Nurmi, J.-E. (1999). *Task-value scale for children*. Unpublished task.

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.* https://doi.org/10.1016/j.learninstruc.2005.04.009

Parent-Reported Children's Experience of Mastery in Math

Parents were asked to evaluate their child's experience of mastery regarding math using two items: "My child thinks that they are good at math" and "My child thinks that math tasks are easy." Parents responded using a 5-point scale: (1 = *not at all*, 5 = *all the time*).

Reference:

Developed in the project.

Parent-Reported Children's Self-Efficacy in Math

Parents were asked to evaluate their child's self-efficacy in math using two items: "My child believes they are able to do math tasks" and "My child believes they are able to learn new math skills." Parents responded using a 5-point scale: (1 = *not at all*, 5 = *all the time*).

Reference:

Developed in the project.

Parent-Reported Children's Math Anxiety

Parents were asked to rate whether their child feels nervous or anxious about mathematics. The measure consisted of 11 statements adapted from Francis et al. (2020). Item examples: "My child worries of not being good at mathematics"; "My child sleeps badly because (they are worried about mathematics" and "My child worries about making mistakes in mathematics." Parents responded using a 4-point scale (1 = *not at all*, 4 = *all the time*).

Reference:

Francis, D., Nation, K., & McArthur, G. (2020). *The Macquarie Oxford Reading Anxiety Test-Parents (MoRAT-P).* Motif.

Home Numeracy Environment (HNE) and Home Literacy Environment (HLE)

The Frequency of Children's Activities in Digital Environments

Parents were asked to evaluate the frequency of their child's activities in digital environments. There were five items, such as: "Child watches shows, series or movies (e.g., on television, streaming services, YouTube)" and "Child uses a computer/mobile device to write or look for information." Parents responded using a 5-point scale (1 = *not at all or rarely*, 5 = *many times a day*).

Reference:

Developed in the project.

Children's Home Math Environment: Informal Activities

Parents were asked to evaluate how often they do math-related activities at home with their child. The measure included eight items. It was based on the measurement in Hart et al. (2016) where 48 statements were used. For example, the following statements were included: "Card games are played with the child"; "Ingredients are measured with the child while cooking" and "The child is taught math." Parents responded using a 6-point scale (1 = *not at all*, 6 = *almost daily or daily*).

Reference:

Hart, S. A., Ganley, C. M., & Purpura, D. J. (2016). Understanding the home math environment and its role in predicting parent report of children's math skills. *PloS One, 11*(12), e0168227. https://doi.org/10.1371/journal.pone.0168227

Time Spent with a Child

Parents were asked to evaluate how much time they approximately spend with their child in a day, including all the activities they do together, such as eating, grocery shopping, and playing. They were instructed to calculate the time the child is awake and to also include the time when there are also other people present. The parents were asked to exclude the time their child is at school, at hobbies or doing some other activities the parent is not involved in. The parents responded to two open-ended items: Time spent with a child "on weekdays – hours/day": ____h, and Time spent with a child "on weekends/holidays – hours/day": ____h.

Reference:

Developed in the project.

Children's Homework and Parental Involvement

Children's Homework: Frequency and Amount of Time Spent

Parents were asked to report about the frequency of their child's homework using the following questions: "Does your child get homework…(a) in mathematics; (b) in language and literacy?" They responded to the items using a 5-point scale (1 = *not at all*, 5 = *daily*). In addition, they

were asked to report the amount of time (minutes per day) their child uses for (1) homework in general, (2) homework in language and literacy and (3) homework in mathematics (open-ended questions).

Reference:

Developed in the project.

Children's Homework: Amount of Time the Parents Help with Homework

Parents were asked to report the amount of time (minutes per day) they use to help their child with (1) homework in general, (2) homework in language and literacy and (3) homework in mathematics (open-ended questions).

Reference:

Developed in the project.

Parental Involvement in Children's Homework: Monitoring and Assistance

Parents were asked to report on their involvement in their child's homework with 16 items. Item examples: "How often do you make sure your child has done the homework?"; "How often do you encourage your child to try to do homework alone first?"; "How often do you check your child's homework together with them?"; "How often do you advise your child to look for some extra information for homework from the internet?" and "How often do you solve your child's homework for them because they don't know how to do something?" Parents responded using a 5-point scale (1 = *never*, 5 = *always when having homework*).

The items 1 to 5 are from Tunkkari et al. (2021); Items 6 to 7 were adapted from Silinskas et al. (2015); Rest of the items were adapted from Gonida & Cortina (2014).

References:

Gonida, E. N., & Cortina, K. (2014). Parent involvement in homework: Relations with parent and student achievement-related motivational beliefs and achievement. *British Journal of Educational Psychology*, *84*, 376–396. https://doi.org/10.1111/bjep.12039

Silinskas, G., Kiuru, N., Aunola, K., Lerkkanen, M.-K., & Nurmi, J.-E. (2015). The developmental dynamics of children's academic performance and mothers' homework-related affect and practices. *Developmental Psychology*, *51*(4), 419. https://doi.org/10.1037/a0038908

Tunkkari, M., Aunola, K., Hirvonen, R., Silinskas, G., & Kiuru, N. (2021). The interplay between maternal homework involvement, task-avoidance, and achievement among adolescents. *Journal of Family Psychology*, *35*(7), 863. https://doi.org/10.1037/fam0000686

Parental Homework-Related Affect

Parents were asked to evaluate how they feel when they help or advise their child with homework. Altogether four items were included: "How often do you…feel satisfied / feel helplessness / feel joy / feel frustration/stress?" Parents responded using a 5-point scale (1 = *not at all*, 5 = *to high extent / very often*). The scale is adapted to parents from Pekrun et al.

(2002) and Watson et al. (1988) and was used before in the First Steps Study (Lerkkanen et al., 2006–2016).

References:

Lerkkanen, M.-K., Niemi, P., Poikkeus, A.-M., Poskiparta, E., Siekkinen, M., & Nurmi, J.-E. (2006–2016). *The First Steps Study* [Alkuportaat]. University of Jyväskylä.

Pekrun, R., Goetz, T., Titz, W., & Perry, R. P. (2002). Academic emotions in students' self-regulated learning and achievement: A program of qualitative and quantitative research. *Educational Psychologist, 37,* 91–105. https://doi.org/10.1207/S15326985EP3702_4

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS Scales. *Journal of Personality & Social Psychology, 54,* 1063–1070. https://doi.org/10.1037/0022-3514.54.6.1063

Parental Belief of Children's Emotions of Being Helped with Homework

Parents were asked to evaluate their beliefs about their child's emotions in situations where the child is being helped or advised by parent with homework. Altogether four items were included: "I believe they feel… satisfied / helplessness / joy / frustration/stress." Parents responded using a 5-point scale (1 = *not at all,* 5 = *to a high extent / very often*). The items were developed in the project based on the work in the First Steps project (Lerkkanen et al., 2006-2016). Pekrun et al. (2009), and Watson et al. (1998).

References:

Lerkkanen, M.-K., Niemi, P., Poikkeus, A.-M., Poskiparta, E., Siekkinen, M., & Nurmi, J.-E. (2006–2016). *The First Steps Study* [Alkuportaat]. University of Jyväskylä.

Pekrun, R., Goetz, T., Titz, W., & Perry, R. P. (2002). Academic emotions in students' self-regulated learning and achievement: A program of qualitative and quantitative research. *Educational Psychologist, 37,* 91–105. https://doi.org/10.1207/S15326985EP3702_4

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS Scales. *Journal of Personality & Social Psychology, 54,* 1063–1070. https://doi.org/10.1037/0022-3514.54.6.1063

Parents' Thoughts about Their Own Role with Their Children's School

Parents were asked to evaluate their thoughts about their own role in their child's school using six items adapted from Walker et al. (2005), such as the following: "I think my role as a guardian is…to help my child with homework"; "…to regularly discuss with my child's teacher" and "…to know what happens in my child's school." Parents responded to the items using a 6-point scale (1 = *strongly disagree,* 6 = *strongly agree*).

Reference:

Walker, J. M., Wilkins, A. S., Dallaire, J. R., Sandler, H. M., & Hoover-Dempsey, K. V. (2005). Parental involvement: Model revision through scale development. *The Elementary School Journal, 106*(2), 85–104. https://doi.org/10.1086/499193

Parents' Self-Efficacy Beliefs on Helping Their Child with Homework

Parents were asked to rate whether they believed they were able to help their child with school. The measure included seven items, such as the following: "I know how I can help my child to succeed in school"; "Even if I try really hard, I'm not able to advise my child with mathematics as well as I'm able to with other homework" and "I have enough time to help my child with homework / school-related things." Parents responded using a 6-point scale (1 = *strongly disagree*, 6 = *strongly agree*). The items were adapted from Hoover-Dempsey & Sandler (1995), O'Sullivan, et al. (2014), Walker, et al. (2005).

References:

Adapted from:

Hoover-Dempsey, K. V., & Sandler, H. M. (1995). Parental involvement in children's education: Why does it make a difference? *Teachers College Record, 97*(2), 310–331. https://doi.org/10.1177/016146819509700202

O'Sullivan, R. H., Chen, Y.-C., & Fish, M. C. (2014). Parental mathematics homework involvement of low-income families [TM1] with middle school students. *School Community Journal*, *24*(2), 165–188. https://psycnet.apa.org/record/2014-56242-008

Walker, J. M., Wilkins, A. S., Dallaire, J. R., Sandler, H. M., & Hoover-Dempsey, K. V. (2005). Parental involvement: Model revision through scale development. *The Elementary School Journal*, *106*(2), 85–104. https://doi.org/10.1086/499193

Children Asking for Help with Homework

Parents were asked to rate the statements regarding whether their child asks their parents to help them with their homework. The items were: "The child asks me to help them…with language and literacy homework / with mathematics homework." Parents responded using a 5-point scale (1 = *never*, 5 = *always when they have homework*).

Reference:

Developed in the project.

Parents' Own Reading and Math Skills and Experiences

Parents' Math Anxiety

Parents were asked to rate the level of their math anxiety with one item: "How anxious are you about math?" Parents responded using a 10-point scale (1 = *not at all*, 10 = *to a high extent*).

References:

Ashcraft, M. H. (2002). Math Anxiety: Personal, Educational, and Cognitive Consequences. Current Directions in Psychological Science, 11(5), 181–185. https://doi.org/10.1111/1467-8721.00196

Núñez-Peña, M. I., Guilera, G., & Suárez-Pellicioni, M. (2014). The Single-Item Math Anxiety Scale: An Alternative Way of Measuring Mathematical Anxiety. Journal of Psychoeducational Assessment, 32(4), 306–317. https://doi.org/10.1177/0734282913508528

Parents' Reading Task Values of Interest and Importance

Parental reading task values of interest and importance were assessed with two items (Aunola & Nurmi, 1999; Nurmi & Aunola, 2005). They were asked to rate whether they like reading with the item: "How much do you like… reading?" Parents responded to the items using a 5-point scale (1 = *not at all*, 5 = *to a high extent*). Parents were asked to rate whether they think reading skills are important in their everyday life: "Are strong reading skills important in your everyday life?" Parents responded to the items using a 5-point scale (1 = *not at all*, 5 = *very important*).

References:

Aunola, K. & Nurmi, J.-E. (1999). *Task-value scale for children*. Unpublished task.

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.* https://doi.org/10.1016/j.learninstruc.2005.04.009

Parents' Math Task Values of Interest and Importance

Parental math task values of interest and importance were assessed with two items (Aunola & Nurmi, 1999; Nurmi & Aunola, 2005). They were asked to rate whether they like math with the item: "How much do you like… math?" Parents responded to the items using a 5-point scale (1 = *not at all*, 5 = *to a high extent*). Parents were asked to rate whether they think math skills are important in their everyday life: "Are strong math skills important in your everyday life?" Parents responded to the items using a 5-point scale (1 = *not at all*, 5 = *very important*).

References:

Aunola, K. & Nurmi, J.-E. (1999). *Task-value scale for children*. Unpublished task.

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.* https://doi.org/10.1016/j.learninstruc.2005.04.009

Parents' Valence Towards School

Parents were asked to report on their experiences about school with six items (Walker et al., 2005). They were asked to rate how well each of the items matched their school experiences when they were a student using a 6-point Likert scale:

1. My school: disliked 1 2 3 4 5 6 liked
2. My teachers: were mean 1 2 3 4 5 6 were nice
3. My teachers: ignored me 1 2 3 4 5 6 cared about me
4. My school experience: bad 1 2 3 4 5 6 good
5. I felt like: An outsider 1 2 3 4 5 6 I belonged
6. My overall experience: failure 1 2 3 4 5 6 success

Reference:

Walker, J. M., Wilkins, A. S., Dallaire, J. R., Sandler, H. M., & Hoover-Dempsey, K. V. (2005). Parental involvement: Model revision through scale development. *The Elementary School Journal*, *106*(2), 85–104. https://doi.org/10.1086/499193

Parents' Anxiety

Parents were asked to evaluate their feelings of anxiety with three items (Viikki & Leinonen, 2015). They were instructed to think of the previous six months and respond to the following items: "I have been remarkably worried or anxious about several things in everyday life at work and home"; "I have been worried or anxious most of the time" and "It has been challenging to control worrying or ruminating or it has distracted my ability to concentrate." Parents responded using a 5-point scale (1 = *applies poorly,* 5 = *applies very well*).

Reference:

Viikki, M., & Leinonen, E. (2015). Ahdistuneisuushäiriöiden diagnostiikka ja lääkehoito perusterveydenhuollossa. *Aikakauskirja Duodecim, 131(6),* 583–590 https://www.duodecimlehti.fi/lehti/2015/6/duo12168

Time Spent: The Use of Different Kind of Media and Devices

The parents were asked to evaluate how much their child uses time in a typical day with different kind of medias or devices, either alone or with someone else. The items were adapted from Ribner & McHarg (2021). The same seven items were asked, first concerning the weekdays and next concerning weekends, comprising altogether 14 items. The items were, for example: "Television (also streaming services, YouTube, DVDs etc. while watching on TV)," "Computer" and "Traditional books or magazines." The parents responded using a 6-point scale (1 = *not at all,* 6 = *5 hours or more*).

Reference:

Ribner, A. D., & McHarg, G. (2021). Screens across the pond: findings from longitudinal screen time research in the US and UK. *Infant Behavior and Development, 63,* 101551. https://doi.org/10.1016/j.infbeh.2021.101551

<u>In addition, the following measure was included in the parental questionnaire for those parents who had not completed the parental questionnaire in Grade 1 or Grade 2:</u>

Adult Math History Questionnaire – Short Version

Parents' math history was assessed with six questions that were adapted from three different measures of adult math/arithmetic history questionnaires that were originally based on the classic Adult Reading History Questionnaire (Lefly & Pennington, 2000). Out of the six items used in this research, five are shared with the AMHQ (Nguyen et al., 2022), AAHQ (Sury & Gaab, 2020), and Stefansson et al.'s (2006) scales. The last question, "Did you experience any difficulties learning the multiplication table in elementary school?", was included in Stefansson et al.'s (2006) measure. The items were as follows: "Did you experience any difficulties in learning math in elementary school?"; "How do you rate your math skills now compared to people your age with a comparable education level?"; "What is your current attitude towards math?"; "If you compare your math skills to people of your education level and age, how would you situate yourself?" and "How much help did you need with math in primary school?" The items were rated on a 5-point scale that varied from item to item, but lower scores indicate less difficulty.

References:

Lefly, D. L., & Pennington, B. F. (2000). Reliability and validity of the adult reading history questionnaire. *Journal of Learning Disabilities*, *33*(3), 286–296. https://doi.org/10.1177/002221940003300306

Nguyen, T. Q., Martinez-Lincoln, A., & Cutting, L. E. (2022). Tracking familial history of reading and math difficulties in children's academic outcomes. *Frontiers in Psychology*, *12*, 710380. https://doi.org/10.3389/fpsyg.2021.710380

Stefansson, H., Meyer-Lindenberg, A., Steinberg, S., Magnusdottir, B., Morgen, K., Arnarsdottir, S., Bjornsdottir, G., Bragi Walters, G., Jonsdottir, G. A., Doyle, O. M., Tost, H., Grimm, O., Kristjansdottir, S., Snorrason, H., Davidsdottir, S. R., Gudmundsson, L. J., Jonsson, G. F, Stefansdottir, B., Helgadottir, I., Haraldsson, M., Jonsdottir, B.,... & Tost, H. (2014). CNVs conferring risk of autism or schizophrenia affect cognition in controls. *Nature*, *505*(7483), 361. https://doi.org/10.1038/nature12818

Sury, D., & Gaab, N. (2020). The Adult Arithmetic History Questionnaire. https://psyarxiv.com/zt6ku/

5.6. TCI – Teacher–Child Interaction Quality (Classroom Observations)

5.6.1. TCI – Toddler Phase (T1–T2)

The Classroom Assessment Scoring System (CLASS-Toddler; LaParo et al., 2012) instrument was used to measure the quality of teacher–child interactions. The CLASS-Toddler is an observational tool designed to analyze the interactions between teachers and children aged 15 to 36 months. The CLASS-Toddler assesses eight dimensions across two broad domains: emotional and behavioral support and engaged support for learning. Emotional and behavioral support encompasses the dimensions of positive climate, negative climate, teacher sensitivity, regard for child perspectives, and behavior guidance. The second domain—engaged support for learning—comprises the dimensions of facilitation of learning and development, quality of feedback, and language modeling. Each dimension is scored on a 7-point scale for each 10- to 20-minute observation cycle, based on the behavioral markers provided in the comprehensive observation manual. Scores 1–2 are seen as low, 3–5 as mid-range, and 6–7 as high quality.

Videos (N = 154) derived from the toddler classrooms were treated as individual observation cycles (N = 154). Two trained CLASS-Toddler observers rated the quality of teacher–child interactions, using the CLASS-Toddler, rating independently each cycle for the eight dimensions. The means for each of the two domains were calculated across dimensions at each cycle and across raters and averaged across the entire observation.

Prior to coding, the two observers attended a two-day CLASS-Toddler observer training with a certified trainer. At the end of the training and before proceeding to coding the actual observation data, each observer coded five master-coded videos and passed the reliability test within 1 scale point of at least 80% of the master codes. At the end, a subset of the video-recorded cycles in the current dataset was double coded to ensure reliability between the two observers across the toddler dataset. Double coding covered 20% of the observed cycles. A

previous study has demonstrated the structural validity of the CLASS-Toddler in Finnish toddler classrooms (Cadima et al., 2022).

References:

Cadima, J., Aguiar, C., Guedes, C., Wysłowska, O., Salminen, J., Slot, P., Barata, C., & Lerkkanen, M.-K. (2022). Process Quality in Toddler Classrooms in Four European Countries. *Early Education and Development, 34*(7), 1565–1589. https://doi.org/10.1080/10409289.2022.2139548

La Paro, K. M., Hamre, B. K., & Pianta, R. C. (2012). *Classroom assessment scoring system. Manual, Toddler.* Paul H. Brookes Publishing Company.

5.6.2. TCI–AC – Toddler Phase (T1–T2)

The Activity Setting measure (Cadima et al., 2018) was used to capture the observable characteristics of the individual video recorded activities in toddler classrooms. The measure aims to tap the situation specific activity characteristics (AC) and it is intended to be used to compliment the global CLASS-scoring of teacher–child interaction quality (TCI), made from the same videos. The Activity Setting measure comprises the following domains: *Adult Involvement*, *Activity Content*, *Materials*, *Social Grouping*, *Engagement with Materials*, *Open-ended*, and *Location of the Activity*. The measure also includes an open question for a small description of the activity taking place (e.g., Observer, Date, Number of children and adults present, Number of children participating in the same activity etc.). The seven domains were separately coded by two trained coders, from the first 15 minutes of each video recording. In cases the video lasted for less than 15 minutes, they coded the entire video.

Adult Involvement represents the role taken by the adult during the activity. This domain was coded each 30-seconds interval, for the first 15 minutes observed in each video. Coding was based on five categories: (a) Non-Involved, (b) Passive, (c) Active-Responsive, (d) Active-Facilitator, and I Active-Directive. Categories were mutually exclusive.

Content/Themes represents the curriculum or pedagogical content of the activity. The observer registered the categories of the content of the activity (e.g., language and literacy, mathematics, arts, etc.), for the 15-minute video. Categories were not mutually exclusive, that is, several categories could be chosen for individual 15-minute video. However, it was preferable to code only the main content.

Materials category represents the materials available during the activity. First, the observer registered the content category (e.g., language and literacy, mathematics, arts etc.) of the materials used during the full 15-minute video. Categories were not mutually exclusive, that is, several categories could be chosen. However, it was preferable to code only the main content. Second, the observer registered the proportion (%) of children using materials in successive 30-seconds intervals according to four levels: 0–24% (of videotaped children are using hands-on materials), 25–49%, 50–74%, 75–100%.

Social Grouping represents the type of grouping predominant in the activity. Each 15-minute video was categorized as large group, small group, pairs/individual, or free. To be coded to a specific category, the social grouping must be observable for more than 50% of the video length. The categories were mutually exclusive.

Open-Ended represents the open-endedness vs. product-orientation of the process or outcome. This domain was only applicable when the activity content was arts-related or when it is expected that children produce something (i.e., a product or a choreography). The categories were mutually exclusive.

Location of the Activity represents the location where the activity occurs (e.g., classroom, atelier, playroom, outdoor etc.). One or more of the categories could be selected, the categories were not mutually exclusive.

The Activity setting training was international and was provided for coders from four countries. One Finnish coder attended this two-day live training. After the training she was required to pass a reliability test by at least 80% agreement in each of the seven domains of the Activity Setting measure across five videos. She then personally trained another Finnish coder, who also took and successfully passed the reliability test, based on the same five videos as the other coder. Eventually, in the dataset 25% of the videos were double coded and Kappa and weighted Kappa scores were computed to indicate the reliability between two raters.

The measure has been used in a study by Cadima et al. (2023).

References:

Cadima, J., Aguiar, C., Guedes, C., Wysłowska, O., Salminen, J., Slot, P., Barata, C., & Lerkkanen, M.-K. (2023). Process quality in toddler classrooms in four European countries. *Early Education and Development, 34*(7), 1565–1589. https://doi.org/10.1080/10409289.2022.2139548

Cadima, J., Guedes, C., Aguiar, T., Slot, P., Wyslowska, O., Salminen, J., & Lerkkanen, M.-K. (2018). *Activity setting measure*. Unpublished manual. Portuguese Foundation for Science and Technology (grant PTDC/MHC-CED/5913/2014, research project "Quality Matters" University of Porto, Portugal).

5.6.3. TCI – Preschool Phase (T3–T4)

The Classroom Assessment Scoring System (CLASS-Pre-K, Pianta et al., 2008) instrument was used to measure the quality of teacher–child interactions in the classrooms of 5- to 6-year-old children. The CLASS Pre-K measures interaction in 10 dimensions of three domains: emotional support, classroom organization and instructional support. The domain of emotional support contains four dimensions: positive climate, negative climate (reversed for analysis), teacher sensitivity, and regard for student perspectives. The domain of classroom organization contains three dimensions: behavior management, productivity, and instructional learning formats. The domain of instructional support includes three dimensions: concept development, quality of feedback, and language modeling. Each dimension is scored on a 7-point scale for each 10- to 20-minute observation cycle, based on the behavioral markers provided in the comprehensive observation manual. Scores 1–2 are seen as low, 3–5 as mid-range, and 6–7 as high quality.

Videos (N = 187) derived from the preschool classrooms were divided into 10- to 20-minute observation cycles (N = 256). Three trained CLASS Pre-K observers rated the cycles according to the 10 CLASS Pre-K dimensions. The means for each of the three domains were calculated across dimensions at each cycle and across raters and averaged across the entire observation.

Prior to coding, the three observers attended a two-day CLASS-Pre-K training session with a certified trainer. At the end of the training and before proceeding to coding the actual observation data, each observer coded five master-coded videos and passed the reliability test within 1 scale point of at least 80% of the master codes. At the end, a subset of the video-recorded cycles in the current dataset was double coded to ensure reliability between the two observers across the preschool dataset. Double coding covered 20% of the observed cycles. A previous study has validated the CLASS-Pre-K in Finnish pre-primary classrooms (Pakarinen et al., 2010).

References:

Pakarinen, E., Lerkkanen, M.-K., Poikkeus, A.-M., Kiuru, N., Siekkinen, M., Rasku-Puttonen, H., & Nurmi, J.-E. (2010). A validation of the classroom assessment scoring system in Finnish kindergartens. *Early Education & Development, 21*(1), 95–124. https://doi.org/10.1080/10409280902858764

Pianta, R. C., K. M. La Paro, & Hamre, B. K. (2008). *Classroom Assessment Scoring System. Manual, Pre-K.* Paul H. Brookes Publishing Company.

5.6.4. TCI – School Phase: Grade 1 (T5)

The Classroom Assessment Scoring System (CLASS-K-3, Pianta et al., 2008) instrument was used to measure the quality of teacher–child interactions in the Grade 1 classrooms. The CLASS K-3 measures interaction in 10 dimensions of three domains: emotional support, classroom organization and instructional support. The domain of emotional support contains four dimensions: positive climate, negative climate (reversed for analysis), teacher sensitivity, and regard for student perspectives. The domain of classroom organization contains three dimensions: behavior management, productivity, and instructional learning formats. The domain of instructional support includes three dimensions: concept development, quality of feedback, and language modeling. Each dimension is scored on a 7-point scale, based on the behavioral markers provided in the comprehensive observation manual. Scores 1–2 are seen as low, 3–5 as mid-range, and 6–7 as high quality.

Each video recording (N = 71) from Grade 1 classrooms was divided into 10- to 20-minute observation cycles (N = 203). Three trained CLASS K-3 observers rated each cycle according to the 10 CLASS K-3 dimensions. The means for each of the three domains were calculated across dimensions at each cycle and across raters and averaged across the entire observation day.

Prior to coding, the three observers attended a two-day CLASS K-3 training session with a certified trainer in October 2021. At the end of the training and before proceeding to coding the actual observation data, each observer coded five master-coded videos and passed the reliability test within 1 scale point of at least 80% of the master codes. At the end, a subset of the video-recorded cycles in the current dataset was double coded to ensure reliability between the two observers across the Grade 1 dataset. Double coding covered 20% of the observed cycles. CLASS K-3 has been reliably used also in the Finnish samples (e.g., Soininen et al., 2023).

References:

Pianta, R. C., K. M. La Paro, & Hamre, B. K. (2008). *Classroom Assessment Scoring System. Manual, K-3.* Paul H. Brookes Publishing Company.

Soininen, V., Pakarinen, E., & Lerkkanen, M.-K. (2023). Reciprocal associations among teacher–child interactions, teachers' work engagement, and children's social competence. *Journal of Applied Developmental Psychology, 85*, 101508. https://doi.org/10.1016/j.appdev.2022.101508

5.6.5. TCI – School Phase: Grade 2 (T6)

The Classroom Assessment Scoring System (CLASS-K-3, Pianta et al., 2008) instrument was used to measure the quality of teacher–child interactions in the Grade 2 classrooms. The CLASS K-3 measures interaction in 10 dimensions of three domains: emotional support, classroom organization and instructional support. The domain of emotional support contains four dimensions: positive climate, negative climate (reversed for analysis), teacher sensitivity, and regard for student perspectives. The domain of classroom organization contains three dimensions: behavior management, productivity, and instructional learning formats. The domain of instructional support includes three dimensions: concept development, quality of feedback, and language modeling. Each dimension is scored on a 7-point scale, based on the behavioral markers provided in the comprehensive observation manual. Scores 1–2 are seen as low, 3–5 as mid-range, and 6–7 as high quality.

Each video recording (N = 70) from Grade 2 classrooms was divided into 10- to 20-minute observation cycles (N = 211). Two trained CLASS K-3 observers rated each cycle according to the 10 CLASS K-3 dimensions. The means for each of the three domains were calculated across dimensions at each cycle and across raters and averaged across the entire observation day.

The observers who coded the Grade 2 data, had attended the official two-day CLASS K-3 training session with a certified trainer in October 2021 and passed the training reliability test within 1 scale point of at least 80% of the master codes. Before starting the Grade 2 CLASS coding, both observers had renewed their yearly CLASS observation certificate to ensure their alignment with the CLASS coding protocol.

At the end, a subset of the video-recorded cycles in the current dataset was double coded to ensure reliability between the two observers across the Grade 2 dataset. Double coding covered 20% of the observed cycles. CLASS K-3 has been reliably used also in the Finnish samples (e.g., Soininen et al., 2023).

References:

Pianta, R. C., K. M. La Paro, & Hamre, B. K. (2008). *Classroom Assessment Scoring System. Manual, K-3.* Paul H. Brookes Publishing Company.

Soininen, V., Pakarinen, E., & Lerkkanen, M.-K. (2023). Reciprocal associations among teacher–child interactions, teachers' work engagement, and children's social competence. *Journal of Applied Developmental Psychology, 85*, 101508. https://doi.org/10.1016/j.appdev.2022.101508

5.6.6. TCI – School Phase: Grade 3 (T7)

The Classroom Assessment Scoring System (CLASS-K-3, Pianta et al., 2008) instrument was used to measure the quality of teacher–child interactions in the Grade 3 classrooms. The

CLASS K-3 measures interaction in 10 dimensions of three domains: emotional support, classroom organization and instructional support. The domain of emotional support contains four dimensions: positive climate, negative climate (reversed for analysis), teacher sensitivity, and regard for student perspectives. The domain of classroom organization contains three dimensions: behavior management, productivity, and instructional learning formats. The domain of instructional support includes three dimensions: concept development, quality of feedback, and language modeling. Each dimension is scored on a 7-point scale, based on the behavioral markers provided in the comprehensive observation manual. Scores 1–2 are seen as low, 3–5 as mid-range, and 6–7 as high quality.

Each video recording (N = 77) from Grade 3 classrooms was divided into 10- to 20-minute observation cycles (N = 226). Two trained CLASS K-3 observers independently rated each cycle according to the 10 CLASS K-3 dimensions. The means for each of the three domains were calculated across dimensions at each cycle and across raters and averaged across the entire observation day.

The observers who coded the Grade 3 data were the same individuals who had coded the Grades 1 and 2 observation datasets. They had both attended the official two-day CLASS K-3 training session with a certified trainer in October 2021 and passed the training reliability test within 1 scale point of at least 80% of the master codes. Before starting the Grade 3 CLASS coding, both observers had again renewed their yearly CLASS observation certificate to ensure their alignment with the CLASS coding protocol.

At the end, a subset of the video-recorded cycles in the current dataset was double coded to ensure reliability between the two observers across the Grade 3 dataset. Double coding covered 20% of the observed cycles. CLASS K-3 has been reliably used also in the Finnish samples (e.g., Soininen et al., 2023).

References:

Pianta, R. C., K. M. La Paro, & Hamre, B. K. (2008). *Classroom Assessment Scoring System. Manual, K-3.* Paul H. Brookes Publishing Company.

Soininen, V., Pakarinen, E., & Lerkkanen, M.-K. (2023). Reciprocal associations among teacher–child interactions, teachers' work engagement, and children's social competence. *Journal of Applied Developmental Psychology, 85,* 101508. https://doi.org/10.1016/j.appdev.2022.101508

5.7. TQ – Teachers' Questionnaires

Teachers' questionnaires were sent in T2 (2.9 years), T4 (5.9 years), T5 (Grade 1), T6 (Grade 2) and T7 (Grade 3). Figures 12–16 provide an overview of the measures grouped in five domains: 1. Teachers' background characteristics, 2. Teachers' occupational well-being, 3. Teachers' self-efficacy, 4. Curriculum activities and curriculum goals in ECEC, and 5. Teaching and teacherhood. The measures are described in detail below the figures.

FIGURE 12 An overview of the **teachers' background characteristics**.

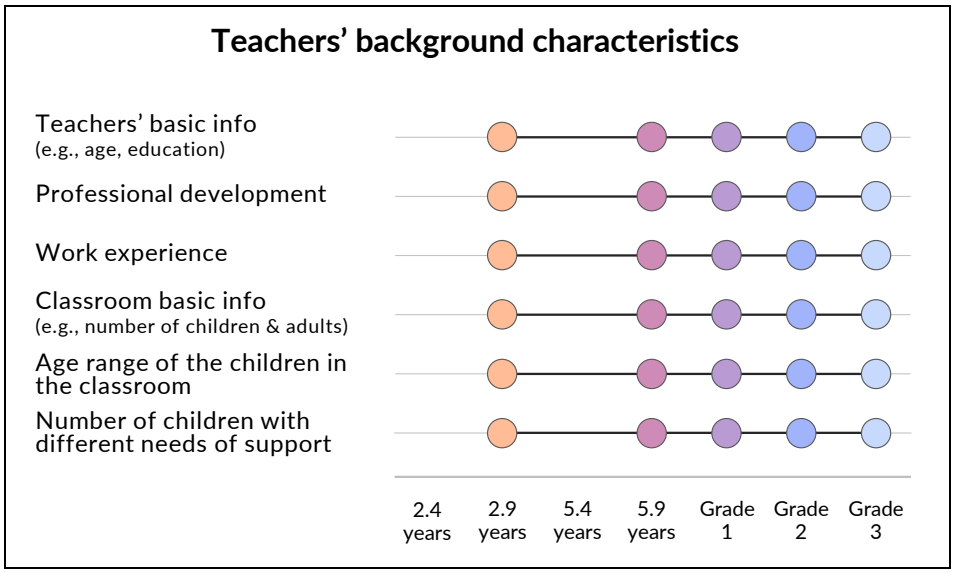
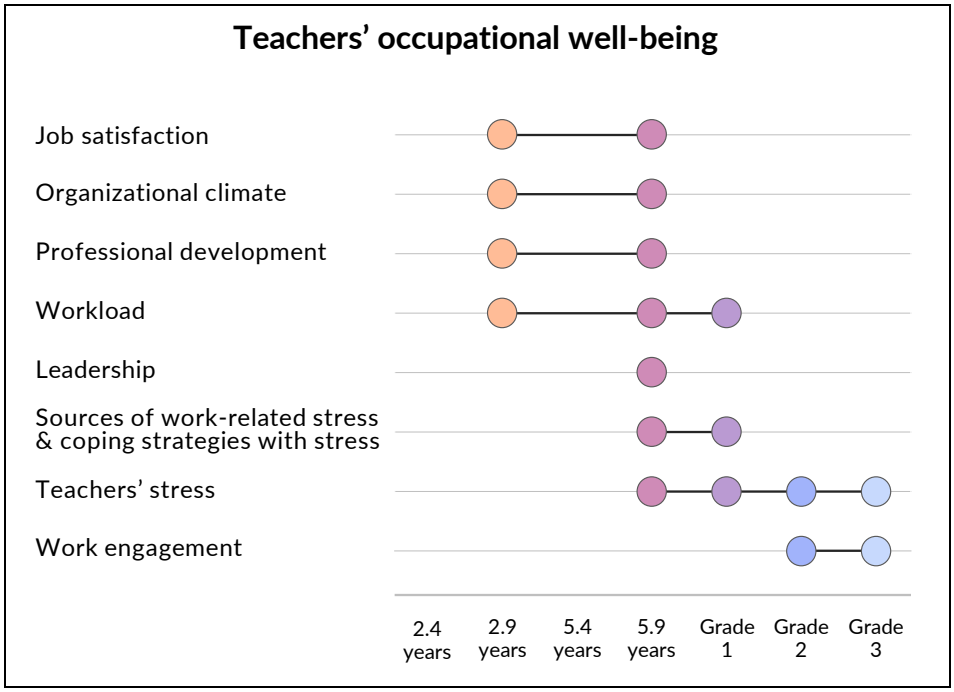


FIGURE 13 An overview of the **teachers' occupational well-being**.



T1–T2 T3–T4 T5 T6 T7

FIGURE 14 An overview of the **teachers' self-efficacy**.

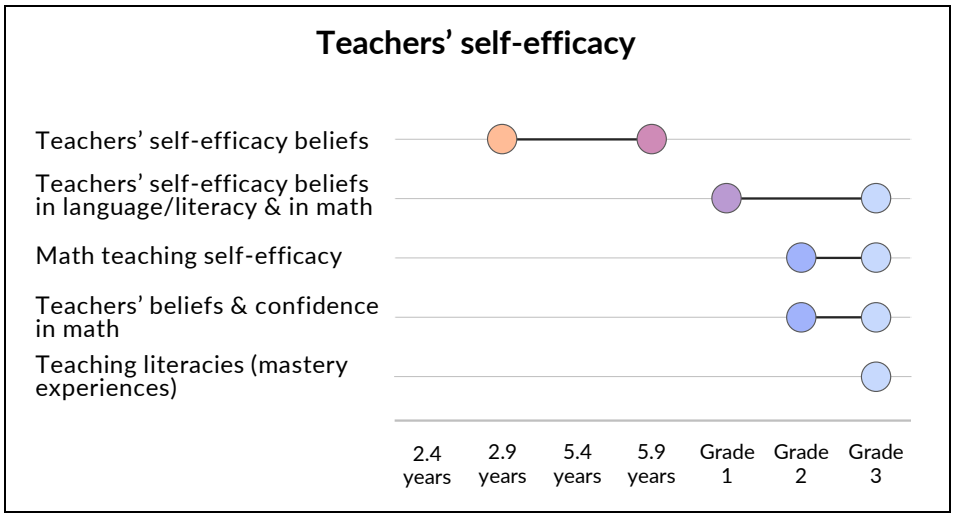


FIGURE 15 An overview of **curriculum activities and curriculum goals in ECEC**.

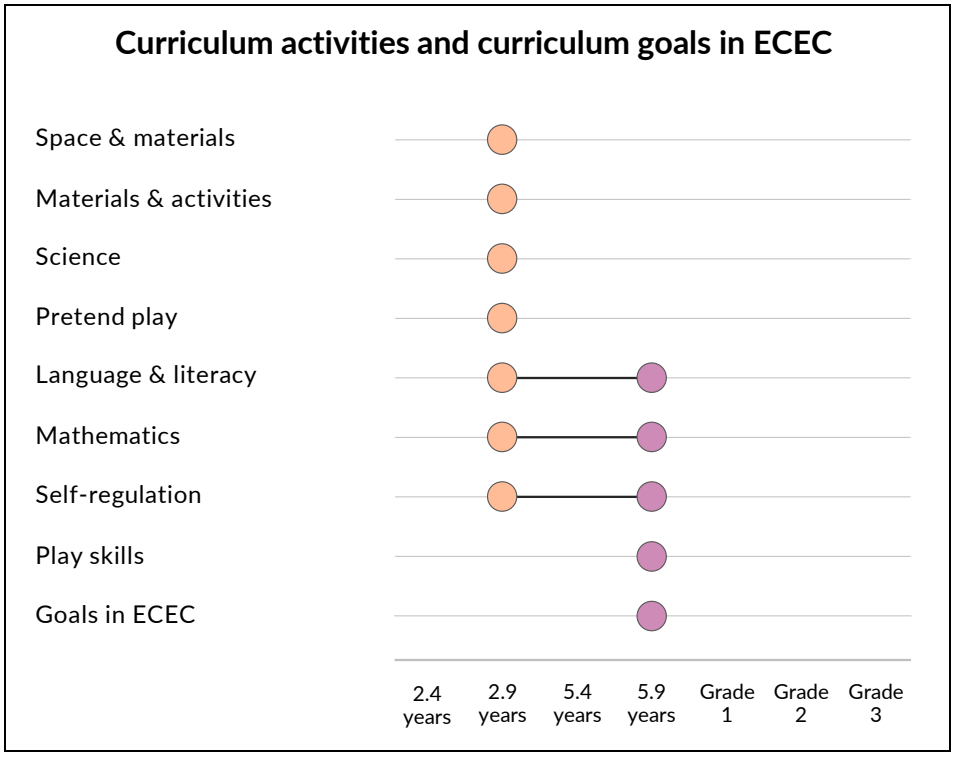
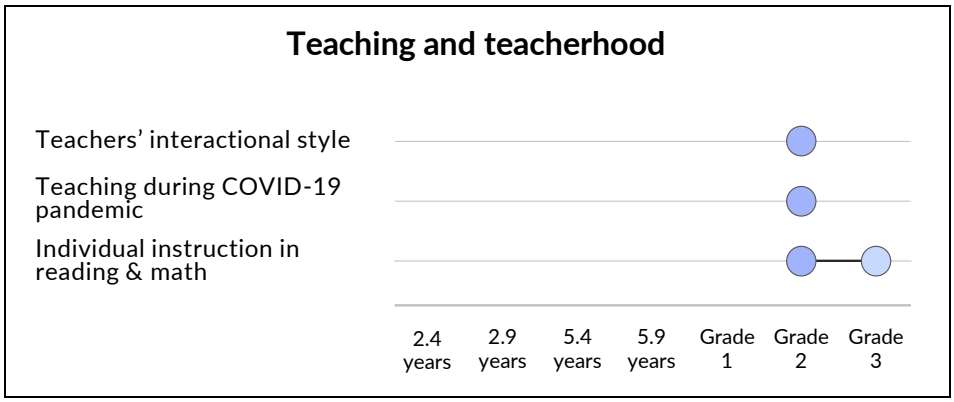


FIGURE 16 An overview of **teaching and teacherhood**.



5.7.1. TQ – Toddler Phase (T1–T2)

TEACHERS' BACKGROUND CHARACTERISTICS

Teacher Characteristics

- Gender: (1) Female; (2) Male
- Year of birth: ____
- Basic education to work in early childhood education? (1) College-level education [opistoasteinen]; (2) Vocational education; (3) University of Applied Sciences / University education
- Do you have a higher education degree? (1) Bachelor's degree; (2) Master's degree; (3) Doctoral degree; (4) Other, what: ____
- Have you participated in professional development over the past two years? (1) No; (2) Yes. If yes:
  - PD1 The content and duration
  - PD2 The content and duration
  - PD3 The content and duration
- Type of employment relationship? (1) Temporary; (2) Permanent; (3) Working as a substitute; (4) Other, what: ____
- Working hours per week?
- Working days per week?
- Work experience in ECEC? ____years or ____months
- Work history in the current ECEC center? ____years or ____months

Classroom Characteristics

- Number of children in the classroom?
- Number of staff in the classroom?
- Is the group work organized in the pair work principle? (1) Yes; (2) No
- Ages of the children in the group? Number of children who are:
  - Less than one year old
  - 1-year-olds
  - 2-year-olds

  - 3-year-olds
  - 4-year-olds
  - 5-year-olds
  - 6-year-olds
- For how long do the children usually stay in your classroom? (stability and change)
- Group composition A: (1) Mainly Finnish children (over 80%); (2) Around half of the children are Finnish and half have an immigrant background; (3) Mainly children with an immigrant background (over 80%)
- Group composition B: Number of children:
  - …with a physical or mental handicap;
  - …with developmental or behavioral challenges;
  - …with developmental delays compared to their peers;
  - …with problem behaviors;
  - …who do not speak the same language with the majority of the population;
  - …who are gifted children or children who are ahead of others in their development.

Teachers' Occupational Well-Being

Job Satisfaction

Job satisfaction was rated by teachers through 18 items and teachers rated their responses on a 5-point scale (1 = *never*, 5 = *always*). Fifteen items were derived from an existing scale used in previous research (pre-COOL Consortium, 2012). Items were, for example: "On the whole, I find my work very meaningful" and "Together as team, we will overcome difficult challenges." Three items were derived from the Parental Stress Inventory (Gerris et al. 1993). The modification involved changing the context from home to ECEC. The three items tapped feelings of stress in ECEC teaching and powerlessness in handling teacher–child situations (e.g., "I have a lot more problems in guiding the children than I expected"). The scale has been used in prior research (e.g., Maio et al., 2022).

References:

Gerris, J. M., Vermulst, A., van Boxtel, D., Janssens, J., van Zutphen, R., & Felling, A. (1993). *Parenting in Dutch families*. University of Nijmegen. Institute of Family Studies.

Maio, R., Guichard, S., & Cadima, J. (2022). In what conditions are intercultural practices implemented in disadvantaged and diverse settings in Portugal? Associations with professional and organization-related variables. *Social Psychology of Education, 25*(2–3), 509–534. https://doi.org/10.1007/s11218-022-09687-6

Pre-COOL Cohortonderzoek (2012). *Technisch rapport tweejarigen onderzoek, eerste meting 2010–2011* [Pre-COOL Cohort Study. Technical report two-year-olds' cohort, first measurement wave 2010–2011]. Kohnstamm Instituut.

Organizational Climate

Organizational climate was rated by teachers using the short form of the Early Childhood Work Environment Survey (Bloom, 2010), which tapped aspects of collegiality, professional growth, supervisor support, clarity, reward system, decision-making, goal consensus, task orientation,

physical setting, and innovativeness. The scale consisted of 20 items, such as: "Staff are friendly and trust one another"; "Morale is high. There is good team spirit" and "Staff are encouraged to learn new skills and competencies." Teachers rated organizational climate on a 6-point Likert scale (1 = *never*, 6 = *always*). The scale has been used in prior research (Slot et al., 2016).

References:

Bloom, P. J. (2010). *Measuring work attitudes in the early childhood setting. Technical manual for the Early Childhood Job Satisfaction Survey and Early Childhood Work Environment Survey* (2nd ed.). McCormick Center for Early Childhood Leadership.

Slot, P., Cadima, J., Salminen, J., Pastori, G., & Lerkkanen, M.-K. (2016). *Multiple case study in seven European countries regarding culture-sensitive classroom quality assessment.* WP2.3 Curriculum and quality analysis impact review. CARE project. Utrecht University. https://ecec-care.org/fileadmin/careproject/Publications/reports/CARE_WP2_D2_3_Multiple_Case_study_FINAL_REPORT.pdf

Professional Development

Professional development activities were rated by teachers using an existing scale used in prior research in ECEC provisions (Slot et al., 2015). The scale evaluates the degree to which different strategies of continuous professional development were implemented at the center. Teachers rated how frequently these activities occurred on a 7-point scale (1 = *never*, 7 = *every day*). The scale consisted of seven items. Item examples were: "Having regular staff meetings to discuss the developmental and educational goals of working with young children"; "Discussing children with special developmental and educational needs" and "Using collegial observation and feedback to improve practice, coaching, and team-based reading of professional literature."

Reference:

Slot, P.L., Leseman, P.P.M., Mulder, H., & Verhagen, J. (2015). Associations between structural quality aspects and process quality in Dutch early childhood education and care settings. *Early Childhood Research Quarterly*, *33*, 64–76. https://doi.org/10.1016/j.ecresq.2015.06.001

Workload

Workload experienced by teachers in their current job was assessed with five items derived from the Bergen Burnout Indicator (Näätänen et al., 2003; see also, Salmela-Aro et al., 2011), which includes 15 items. Teachers assessed the items on a 6-point Likert scale (1 = *totally disagree*, 6 = *totally agree*). Items were, for example: "I am snowed under with work" and "I often sleep poorly because of the circumstances at work."

References:

Näätänen, P., Aro, A., Matthiesen, S., & Salmela-Aro, K. (2003). *Bergen Burnout Indicator 15*. Edita.

Salmela-Aro, K., Rantanen, J., Hyvönen, K., Tilleman, K., & Feldt, T. (2011). Bergen Burnout Inventory: reliability and validity among Finnish and Estonian managers.

International Archives of Occupational and Environmental Health, 84, 635–645. https://doi.org/10.1007/s00420-010-0594-3

Teachers' Self-Efficacy

Teachers' Self-Efficacy Beliefs

The teachers' self-efficacy beliefs were measured with the Teachers' Sense of Efficacy Scale (TSES; Tschannen-Moran & Woolfolk Hoy, 2001). The original measure includes 24 items that are assessed on a 5-point scale (1 = *not at all*, 5 = *to a great deal*). The TSES measure includes three 8-item subscales: (1) efficacy for instructional strategies, (2) efficacy for classroom management, and (3) efficacy for student engagement. In the current timepoint, teachers rated their teaching related self-efficacy with 9 out of 24 original TSES items. The items were selected based on the CARE Multiple Case study (Slot et al., 2016), where three items representing each of the TSES subscales were selected. The 9 items were rated on a 5-point Likert scale (1 = *not at all*, 5 = *to a great deal*). The selected items were as follows: (1) efficacy for instructional strategies: "How much can you do to adjust your practices according to the level of individual children?"; "To what extent can you provide an alternative explanation or example when children are confused?" "How well can you provide appropriate challenges for very capable children?"; (2) efficacy for classroom management: "How much can you do to control disruptive behavior in the classroom?" "How well can you respond to defiant children?"; "How much can you do to calm a child who is disruptive or noisy?" (3) efficacy for student engagement: "How much can you do to get through to the most difficult children?"; "How much can you do to support children who need additional support?"; "How much can you assist families in supporting their child's learning and development?".

Reference:

Slot, P., Cadima, J., Salminen, J., Pastori, G., & Lerkkanen, M.-K. (2016). *Multiple case study in seven European countries regarding culture-sensitive classroom quality assessment*. WP2.3 Curriculum and quality analysis impact review. CARE project. Utrecht University. https://ecec-care.org/fileadmin/careproject/Publications/reports/CARE_WP2_D2_3_Multiple_Case_study_FINAL_REPORT.pdf

Tschannen-Moran, M. & Woolfolk Hoy, A. (2001). Teacher efficacy: capturing an elusive construct. *Teaching and Teacher Education 17*, 783–805. https://doi.org/10.1016/S0742-051X(01)00036-1

Curriculum Activities and Curriculum Goals in ECEC

Space and Materials

Teachers assessed the physical spaces in their ECEC classroom from the perspective of having areas, materials and routines for different daily activities (16 items). The items were adapted from the ITERS (Harms et al., 2003) and ECERS (Harms et al., 1998) observation systems by the CARE research team (see Slot et al., 2016). Teachers rated items on a 5-point scale in terms of how well the given statements applied to their classroom (1 = *does not apply*, 5 = *applies very well*). Exemplary items include "There is an area in my classroom that children can use

for reading and looking at books" and "Daily schedule (depicted, e.g., with pictures) is available for children to see."

References:

Harms, T., Clifford, R., & Cryer, D. (1998). *Early childhood environment rating scale–Revised*. Teachers College Press

Harms, T., Cryer, D., & Clifford, R.M. (2003). *Infant/Toddler Environment Rating Scale–Revised*. Teachers College Press.

Slot, P., Cadima, J., Salminen, J., Pastori, G., & Lerkkanen, M.-K. (2016). *Multiple case study in seven European countries regarding culture-sensitive classroom quality assessment*. WP2.3 Curriculum and quality analysis impact review. CARE project. Utrecht University. https://ecec-care.org/fileadmin/careproject/Publications/reports/CARE_WP2_D2_3_Multiple_Case_study_FINAL_REPORT.pdf

Materials and Activities

Teachers assessed the availability of different learning materials in their classrooms that facilitate learning opportunities for the children (17 items). Teachers rated items on a 5-point scale in terms of how well the given statements applied to their classroom (1 = *does not apply*, 5 = *applies very well*). Exemplary items are: "There are many books in my classroom, meeting children's interests and developmental stages" and "There is a variety of materials and props available for role play in my classroom (e.g., dresses, kitchen items, dolls)."

References:

Harms, T., Clifford, R., & Cryer, D. (1998). *Early childhood environment rating scale–Revised*. Teachers College Press

Harms, T., Cryer, D., & Clifford, R.M. (2003). *Infant/Toddler Environment Rating Scale–Revised*. Teachers College Press.

Slot, P., Cadima, J., Salminen, J., Pastori, G., & Lerkkanen, M.-K. (2016). *Multiple case study in seven European countries regarding culture-sensitive classroom quality assessment*. WP2.3 Curriculum and quality analysis impact review. CARE project. Utrecht University. https://ecec-care.org/fileadmin/careproject/Publications/reports/CARE_WP2_D2_3_Multiple_Case_study_FINAL_REPORT.pdf

Provision of Different Types of Curriculum Activities

The different curriculum activity scales (science, language and literacy, math) and scales on pretend play and self-regulation were developed for the CARE Multiple Case study (Slot et al., 2016) based on and adapted from an existing educator questionnaire used in the longitudinal Dutch cohort study pre-COOL (see, Slot et al., 2015). Answers on all scales were rated on a 7-point scale (1 = *never*, 7 = *three or more times a day*). Below is a description of the curriculum activity scales used.

Science activities (7 items) assesses the degree to which teachers use activities, conversations, and play related to science. Example item: "Comparing and discussing different seeds and pits (for instance, that a flower, fruit or tree grows out of this."

Language and literacy activities (18 items) measures the average frequency of different language activities such as singing songs, rhyming and having classroom conversations, along with the average frequency of the provision of activities involving literacy or literacy materials. Example items: "Having elaborate conversations about children's personal experiences, for instance what they did on the weekend" and "Asking the children questions about the content of the story during or after reading the story."

Math activities (12 items) represents the average frequency of different number and math activities, for instance counting and sorting activities, and activities exploring different shapes. Example item: "Counting how many objects you have, for example counting to five and saying 'I have five marbles'."

Pretend play (8 items) represents the degree to which the teacher encourages cognitive distancing, symbolizing and pretend play in children by modelling behavior and encouraging children to participate in symbolic and pretend play. Example item: "I show children how to use an object for something else then intended, for instance driving a wooden block as if it is a car."

Self-regulation (11 items) evaluates the extent to which the teacher uses routines, activities and play to stimulate children's behavioral self-regulation, such as talking about emotions and feelings, supporting them in resolving peer conflicts or playing games involving turn taking. Example item: "When children have a conflict, I let them express their own opinion, so they better understand what the other thinks."

References:

Pre-COOL Cohortonderzoek (2012). *Technisch rapport tweejarigen onderzoek, eerste meting 2010–2011* [Pre-COOL Cohort Study. Technical report two-year-olds' cohort, first measurement wave 2010–2011]. Kohnstamm Instituut.

Slot, P., Cadima, J., Salminen, J., Pastori, G., & Lerkkanen, M.-K. (2016). *Multiple case study in seven European countries regarding culture-sensitive classroom quality assessment.* WP2.3 Curriculum and quality analysis impact review. CARE project. Utrecht University. https://ecec-care.org/fileadmin/careproject/Publications/reports/CARE_WP2_D2_3_Multiple_Case_study_FINAL_REPORT.pdf

Slot, P.L., Leseman, P.P.M., Mulder, H., & Verhagen, J. (2015). Associations between structural quality aspects and process quality in Dutch early childhood education and care settings. *Early Childhood Research Quarterly, 33*, 64–76. https://doi.org/10.1016/j.ecresq.2015.06.001

5.7.2. TQ – Preschool Phase (T3–T4)

TEACHERS' BACKGROUND CHARACTERISTICS

Teacher Characteristics

- Gender: (1) Female; (2) Male; (3) I prefer not to answer
- Year of birth: ____

- Marital status: (1) Married; (2) Divorced; (3) Common law marriage; (4) I live alone; (5) I prefer not to answer.
- How many members are there in your family, currently living in the same household?
- What is the average income of your household (gross income before taxes) (1) Less than €500/month; (2) €500–€1000/month; (3) €1001–€1499/month; (4) €1500–€1999/month; (5) €2000–€2499/month; (6) €2500–2999 €/month; (7) €3000–€4999/month; (8) €5000–€7499/month; (9) €7500–€10,000/month; (10) Over €10,000/month
- What is your basic education to work in early childhood education? (1) College-level education [opistoasteinen]; (2) Vocational education; (3) University of Applied Sciences / University education
- Do you have a higher education degree? (1) Bachelor's degree; (2) Master's degree; (3) Doctoral degree; (4) Other, Other, what: ____
- Have you participated in professional development over the past five years? (1) No; (2) Yes. If yes:
    - PD1 The content and duration
    - PD2 The content and duration
    - PD3 The content and duration
- What is your professional title? (1) Teacher in ECEC; (2) Social Pedagogue in ECEC; (3) Childcarer in ECEC; (4) Special education teacher in ECEC; (5) Other, what: ____
- Type of employment relationship? (1) Temporary; (2) Permanent; (3) Working as a substitute; (4) Other, what: ____
- Working hours per week?
- Working days per week?
- Work experience in ECEC? ____years and/or ____months
- Work history in the current ECEC center? ____years and/or ____months

Classroom Characteristics

- Number of children in your current ECEC center?
- Number of children in the classroom?
- Number of staff in the classroom?
- Is the group work organized in the pair work principle? (1) Yes; (2) No
- Ages of the children in the group? Number of children who are:
    - Less than one year old
    - 1-year-olds
    - 2-year-olds
    - 3-year-olds
    - 4-year-olds
    - 5-year-olds
    - 6-year-olds
- Group composition A: (1) Mainly Finnish children (over 80%); (2) Around half of the children are Finnish and half have an immigrant background; (3) Mainly children with an immigrant background (over 80%)
- Group composition B: Number of children:
    - …with needs for support in learning;
    - …with needs for support in socio-emotional development or behavior;
    - …who do not speak Finnish as their native language;
    - …who are gifted children or children who are ahead of others in their development.

Teachers' Occupational Well-Being

Job Satisfaction

Job satisfaction was rated by teachers through 18 items and teachers rated their responses on a 5-point scale (1 = *never*, 5 = *always*). Fifteen items were derived from an existing scale used in previous research (pre-COOL Consortium, 2012). Items were, for example: "On the whole, I find my work very meaningful" and "Together as team, we will overcome difficult challenges." Three items were derived from the Parental Stress Inventory (Gerris et al. 1993). The modification involved changing the context from home to ECEC. The three items tapped feelings of stress in ECEC teaching and powerlessness in handling teacher–child situations (e.g., "I have a lot more problems in guiding the children than I expected"). The scale has been used in prior research (e.g., Maio et al., 2022).

References:

Gerris, J. M., Vermulst, A., van Boxtel, D., Janssens, J., van Zutphen, R., & Felling, A. (1993). *Parenting in Dutch families.* University of Nijmegen. Institute of Family Studies.

Maio, R., Guichard, S., & Cadima, J. (2022). In what conditions are intercultural practices implemented in disadvantaged and diverse settings in Portugal? Associations with professional and organization-related variables. *Social Psychology of Education, 25*(2–3), 509–534. https://doi.org/10.1007/s11218-022-09687-6

Pre-COOL Cohortonderzoek (2012). *Technisch rapport tweejarigen onderzoek, eerste meting 2010–2011* [Pre-COOL Cohort Study. Technical report two-year-olds' cohort, first measurement wave 2010–2011]. Kohnstamm Instituut.

Organizational Climate

Organizational climate was rated by teachers using the short form of the Early Childhood Work Environment Survey (Bloom, 2010), which tapped aspects of collegiality, professional growth, supervisor support, clarity, reward system, decision-making, goal consensus, task orientation, physical setting, and innovativeness. The scale consisted of 20 items, such as "Staff are friendly and trust one another"; "Morale is high. There is good team spirit" and "Staff are encouraged to learn new skills and competencies." Teachers rated organizational climate on a 6-point Likert scale (1 = *never*, 6 = *always*). The scale has been used in prior research (Slot et al., 2016).

References:

Bloom, P. J. (2010). *Measuring work attitudes in the early childhood setting. Technical manual for the Early Childhood Job Satisfaction Survey and Early Childhood Work Environment Survey* (2nd ed.). McCormick Center for Early Childhood Leadership.

Slot, P., Cadima, J., Salminen, J., Pastori, G., & Lerkkanen, M.-K. (2016). *Multiple case study in seven European countries regarding culture-sensitive classroom quality assessment*. WP2.3 Curriculum and quality analysis impact review. CARE project. Utrecht University. https://ecec-care.org/fileadmin/careproject/Publications/reports/CARE_WP2_D2_3_Multiple_Case_study_FINAL_REPORT.pdf

Professional Development

Professional development activities were rated by teachers using an existing scale used in prior research in ECEC provision (Slot et al., 2015). The scale evaluates the degree to which different strategies of continuous professional development were implemented at the center. Teachers rated how frequently these activities occurred on a 7-point scale (1 = *never*, 7 = *every day*). The scale consisted of seven items. Item examples were: "Having regular staff meetings to discuss the developmental and educational goals of working with young children"; "Discussing children with special developmental and educational needs" and "Using collegial observation and feedback to improve practice, coaching, and team-based reading of professional literature."

Reference:

Slot, P.L., Leseman, P.P.M., Mulder, H., & Verhagen, J. (2015). Associations between structural quality aspects and process quality in Dutch early childhood education and care settings. *Early Childhood Research Quarterly, 33*, 64–76. https://doi.org/10.1016/j.ecresq.2015.06.001

Workload

Workload experienced by teachers in their current job, was assessed with 11 items selected from the Bergen Burnout Indicator (Näätänen et al., 2003; see also Salmela-Aro et al., 2011), which includes 15 items. Items cover three areas of work-related burnout: exhaustion, cynicism, and inadequacy. Teachers assessed the items on a 6-point Likert scale (1 = *totally disagree*, 6 = *totally agree*). Items were, for example, "I am snowed under with work"; "I feel dispirited at work and I think of leaving my job" and "I frequently question the value of my work."

References:

Näätänen, P., Aro, A., Matthiesen, S., & Salmela-Aro, K. (2003). *Bergen Burnout Indicator 15*. Edita.

Salmela-Aro, K., Rantanen, J., Hyvönen, K., Tilleman, K., & Feldt, T. (2011). Bergen Burnout Inventory: reliability and validity among Finnish and Estonian managers. *International Archives of Occupational and Environmental Health, 84*, 635–645. https://doi.org/10.1007/s00420-010-0594-3

Leadership

Teachers assessed the extent to which the leader of the ECEC setting supports the vision and mission of the ECEC center and provides professional support at individual and team levels with Global Transformational Leadership scale (GTL; Carless et al, 2000). The GTL scale includes seven items which comprise following sections: Vision ("Communicates a clear and positive vision of the future"), Staff development ("Treats staff as individuals, supports and encourages their development"), Supportive leadership ("Gives encouragement and recognition to staff"), Empowerment ("Empowers by fostering trust, involvement and co-operation among team members"), Innovative or lateral thinking ("Encourages thinking about problems in new ways and questions assumptions"), Lead by example ("Is clear about their values and practices what they preach") and Charismatic leadership ("Instills pride and respect in others and inspires me by being highly competent"). Teachers were asked to evaluate with

a 5-point scale (1 = *not at all*, 5 = *almost always*) to what extent the leaders in the school acted according to the items listed above.

Reference:

Carless, S. A., Wearing, A. J., & Mann, L. (2000). A short measure of transformational leadership. *Journal of Business and Psychology, 14*, 389–405. https://doi.org/10.1023/A:1022991115523

Sources of Work-Related Stress and Coping Strategies with Stress

Sources of work-related stress were measured by asking them to write down their answers to an open-ended question: "What causes you the most stress and exhaustion at work?" (Elomaa et al., 2020).

Coping strategies were measured by asking them to write down their answers to an open-ended question: "What are your ways of coping with work-related stress and exhaustion?" (Aulèn et al., 2021; Elomaa et al., 2020).

References:

Aulèn, A.-M., Pakarinen, E., Feldt, T., & Lerkkanen, M.-K. (2021). Teacher coping profiles in relation to teacher well-being: A mixed-method approach. *Teaching and Teacher Education, 102*, 103323. https://doi.org/10.1016/j.tate.2021.1033230742-051X

Elomaa, M., Pakarinen, E., Eskelä-Haapanen, S., Halttunen, L., Von Suchodoletz, A. & Lerkkanen, M.-K. (2020). Directors' Stress in Day Care Centers: Related Factors and Coping Strategies. *International Journal of Educational Management*. https://doi.org/10.1108/IJEM-10-2019-0383

Teachers' Stress

Teachers' stress was measured with a question which is part of the Occupational Stress Questionnaire: "Stress means a situation in which a person feels tense, restless, nervous or anxious or is unable to sleep at night because their mind is troubled all the time. Do you feel this kind of stress these days?" (Elo et al., 2003). The single-item stress measure has been verified as acceptable for measuring variance in well-being between different groups (Aulèn et al., 2021; Elo et al., 2003; Pöysä et al., 2021). Teachers rated the item with a 6-point Likert scale (1 = *not at all*, 6 = *very much*).

The degree of teachers work related stress was further inquired with one continuation item ("To what extent do you feel that stress is being caused by your work?"). The question has been developed in TESSI study (Lerkkanen & Pakarinen, 2021) and used in a study by Aulén et al., (2022). Teachers rated the item on a 4-point Likert scale (1 = *not at all*, 4 = *entirely/fully*).

References:

Aulèn, A.-M., Pakarinen, E., Feldt, T., & Lerkkanen, M.-K. (2021). Teacher coping profiles in relation to teacher well-being: A mixed-method approach. *Teacher and Teacher Education, 102*, 103323. https://doi.org/10.1016/j.tate.2021.1033230742-051X

Aulén, A.-M., Pakarinen, E., Feldt, T., Tolvanen, A., & Lerkkanen, M.-K. (2022). Psychological detachment as a mediator between successive days' job stress and

negative affect of teachers. *Frontiers in Education, 7*, 903606. https://doi.org/10.3389/feduc.2022.903606

Elo, A., Leppänen, A., & Jahkola, A. (2003). Validity of a single-item measure of stress symptoms. *Scandinavian Journal of Work, Environment & Health, 29*(6), 444–451. https://doi.org/10.5271/sjweh.752

Lerkkanen, M.-K., & Pakarinen, E. (2021). *Teacher and Student Stress and Interaction in Classroom (TESSI).* V. 31.7.2021. https://doi.org/10.17011/jyx/dataset/77741

Pöysä, S., Pakarinen, E., & Lerkkanen, M.-K. (2021). Patterns of teachers' occupational well-being during the Covid-19 pandemic: Relations to experiences of exhaustion and recovery, and interactional styles of teaching. *Frontiers in Education, 6*, 699785. https://doi.org/10.3389/feduc.2021.699785.

Teachers' Self-Efficacy

Teachers' Self-Efficacy Beliefs

The teachers' self-efficacy beliefs were measured with the Teachers' Sense of Efficacy Scale (TSES; Tschannen-Moran & Woolfolk Hoy, 2001). The measure includes 24 items that are assessed on a 5-point scale (1 = *not at all*, 5 = *to a great deal*). The TSES measure includes three 8-item subscales: (1) efficacy for instructional strategies (e.g., "To what extent can you provide an alternative explanation or example when children are confused?"), (2) efficacy for classroom management (e.g., "How much can you do to control disruptive behavior in the classroom?") and (3) efficacy for student engagement (e.g., "How much can you do to motivate children who show low interest toward activities in ECEC?").

Reference:

Tschannen-Moran, M. & Woolfolk Hoy, A. (2001). Teacher efficacy: capturing an elusive construct. *Teaching and Teacher Education 17*, 783–805. https://doi.org/10.1016/S0742-051X(01)00036-1

Curriculum Activities and Curriculum Goals in ECEC

Provision of Different Types of Curriculum Activities

The different curriculum activity scales (language and literacy, math) and scale on self-regulation were originally developed for the CARE Multiple Case study (Slot et al., 2016) based on and adapted from an existing educator questionnaire used in the longitudinal Dutch cohort study pre-COOL (see, Slot et al., 2015). The language and literacy and math scales were further modified by the VUOKKO research team. Below is a description of the curriculum activity scales used.

Language and Literacy Activities were measured with 24 items. Eleven items were derived from the scale originally developed for the CARE Multiple Case study (Slot et al., 2016) based on and adapted from an existing educator questionnaire used in the longitudinal Dutch cohort study pre-COOL (Slot et al., 2015). Thirteen items were added and modified by the VUOKKO research team. The 11 items cover the average frequency of different language activities, such

as singing songs, rhyming, and having classroom conversations, along with the average frequency of the provision of activities involving literacy or literacy materials. Examples of items are "Having elaborate conversations about children's personal experiences, for instance what they did on the weekend" and "Asking the children questions about the content of the story during or after reading the story." The 13 items cover the early years of literacy practices and supporting the emerging reading skills, such as "Identifying initial phonemes" and "Practicing naming and writing letters." Answers on all items were rated on a 7-point scale (1 = *never*, 7 = *three or more times a day*).

References:

Pre-COOL Cohortonderzoek (2012). *Technisch rapport tweejarigen onderzoek, eerste meting 2010–2011* [Pre-COOL Cohort Study. Technical report two-year-olds' cohort, first measurement wave 2010–2011]. Kohnstamm Instituut.

Slot, P., Cadima, J., Salminen, J., Pastori, G., & Lerkkanen, M.-K. (2016). *Multiple case study in seven European countries regarding culture-sensitive classroom quality assessment.* WP2.3 Curriculum and quality analysis impact review. CARE project. Utrecht University. https://ecec-care.org/fileadmin/careproject/Publications/reports/CARE_WP2_D2_3_Multiple_Case_study_FINAL_REPORT.pdf

Slot, P.L., Leseman, P.P.M., Mulder, H., & Verhagen, J. (2015). Associations between structural quality aspects and process quality in Dutch early childhood education and care settings. *Early Childhood Research Quarterly, 33*, 64–76. https://doi.org/10.1016/j.ecresq.2015.06.001

Math Activities were measured with 24 items. Nine items were derived from the scale originally developed for the CARE Multiple Case study (Slot et al., 2016) based on and adapted from an existing educator questionnaire used in the longitudinal Dutch cohort study pre-COOL (Slot et al., 2015). These represent the average frequency of different number and math activities, for instance counting and sorting activities, and activities exploring different shapes. An example item: "Counting how many objects you have, for example counting to five and saying 'I have five marbles'." Ten items were added and modified by the VUOKKO research team. These 10 items covered practicing central mathematical skills in preschool, such as "Reciting number sequences backwards (8,7,6…)"; "Comparing numbers/quantities below 10 (less, more)" and "Practicing problem-solving." Finally, five items were adapted from the home math environment scale by LeFevre et al. (2009) by changing the context from home to ECEC. The items included, for example, "Playing board games with die or money" and "Using calendars and discussing dates." Answers on all 24 items were rated on a 7-point scale (1 = *never*, 7 = *three or more times a day*).

References:

LeFevre, J. A., Skwarchuk, S. L., Smith-Chant, B. L., Fast, L., Kamawar, D., & Bisanz, J. (2009). Home numeracy experiences and children's math performance in the early school years. *Canadian Journal of Behavioural Science/Revue 144anadienne des sciences du comportement, 41*(2), 55. https://doi.org/10.1037/a0014532

Slot, P., Cadima, J., Salminen, J., Pastori, G., & Lerkkanen, M.-K. (2016). Multiple case study in seven European countries regarding culture-sensitive classroom quality assessment. WP2.3 Curriculum and quality analysis impact review. CARE project. Utrecht University. https://ecec-

care.org/fileadmin/careproject/Publications/reports/CARE_WP2_D2_3_Multiple_Case_study_FINAL_REPORT.pdf

Slot, P.L., Leseman, P.P.M., Mulder, H., & Verhagen, J. (2015). Associations between structural quality aspects and process quality in Dutch early childhood education and care settings. *Early Childhood Research Quarterly, 33*, 64–76. https://doi.org/10.1016/j.ecresq.2015.06.001

Self-Regulation activities were measured with 11 items originally developed for the CARE Multiple Case study (Slot et al., 2016) based on and adapted from an existing educator questionnaire used in the longitudinal Dutch cohort study pre-COOL (Slot et al., 2015). These items evaluate the extent to which the teacher uses routines, activities and play to stimulate children's behavioral self-regulation, such as talking about emotions and feelings, supporting them in resolving peer conflicts or playing games involving turn taking. Example item: "When children have a conflict, I let them express their own opinion, so they better understand what the other thinks." Answers on all items were rated on a 7-point scale (1 = *never*, 7 = *three or more times a day*).

References:

Slot, P., Cadima, J., Salminen, J., Pastori, G., & Lerkkanen, M.-K. (2016*). Multiple case study in seven European countries regarding culture-sensitive classroom quality assessment.* WP2.3 Curriculum and quality analysis impact review. CARE project. Utrecht University. https://ecec-care.org/fileadmin/careproject/Publications/reports/CARE_WP2_D2_3_Multiple_Case_study_FINAL_REPORT.pdf

Slot, P.L., Leseman, P.P.M., Mulder, H., & Verhagen, J. (2015). Associations between structural quality aspects and process quality in Dutch early childhood education and care settings. *Early Childhood Research Quarterly, 33*, 64–76. https://doi.org/10.1016/j.ecresq.2015.06.001

Supporting Children's Play Skills. Teachers were asked to evaluate the ways in which the staff supports children's play skills in their group. Supporting play skills was measured through nine items, which the teacher rated on a 5-point scale (1 = *completely disagree*, 5 = *completely agree*). Items were adapted for the VUOKKO study from the report by national education evaluation center (FINEEC: Repo et al., 2019). The items included "We observe children's play in a systematic way in our classroom" and "Teachers in our classroom engage actively in children's play on a daily basis."

Reference:

Repo, L., Paananen, M., Eskelinen, M., Mattila, V., Lerkkanen, M-K., Gammelgård, L., Ulvinen, J., Marjanen, J., Kivistö, A., & Hjelt, H. (2019). *Varhaiskasvatuksen laatu arjessa. Varhaiskasvatussuunnitelmien toteutuminen päiväkodeissa ja perhepäivähoidossa* [Every-day quality in early childhood education and care – ECEC curriculum implementation at day-care centers and in family day-care] (Julkaisut 15:2019). Kansallinen koulutuksen arviointikeskus [Finnish National Education Evaluation Centre].

Goals in ECEC

Teachers were asked to reflect on their values regarding ECEC through one open ended question: “In your opinion, what is the most central goal for 5- to 6-year-old children in ECEC?”

Reference:

Developed in the study.

5.7.3. TQ – School Phase: Grade 1 (T5)

TEACHERS’ BACKGROUND CHARACTERISTICS

Teacher Characteristics

- Gender: (1) Female; (2) Male; (3) I prefer not to answer
- Age in years: ____
- Basic education (1) Class teacher; (2) Dual qualified class teacher; (3) Subject teacher; (4) Special education teacher; (5) Other, what: ____
- Current occupational title: (1) Class teacher; (2) Special class teacher; (3) Other, what: ____
- Have you participated in professional development over the past five years? (1) No; (2) Yes. If yes:
  - PD1 The content and duration
  - PD2 The content and duration
  - PD3 The content and duration
- Work experience: (a) In school: ____years and/or ____months; (b) Other teaching experience: ____years and/or ____months
- Work history in the current school: ____years and/or ____months
- Type of employment relationships (temporary, permanent, substitute other)

Classroom Characteristics

- Number of additional staff in the classroom: Number: ____ Professional titles: ____
- Availability of additional staff in the classroom: ____hours/week
- Number of students in the classroom: ____students, of which ____boys and ____girls
- Classroom composition: Number of children:
  - …with needs for support in some areas of learning;
  - …with needs for support in socio-emotional development or behavior;
  - …who do not speak Finnish as their native language;
- …gifted children or children who are ahead of others in development.
- Ages of the children in the classroom: (a) 6-year-olds (born in 2014); (b) 7-year-olds (born in 2013); (c) 8-year-olds (born in 2012); (d) 9-year-olds (born in 2011); (e) Other, born in ____

Teachers' Occupational Well-Being

Workload

Workload experienced by teachers in their current job was assessed with 11 items selected from the Bergen Burnout Indicator (Näätänen et al., 2003), which includes 15 items. Items cover three areas of work-related burnout: exhaustion, cynicism, and inadequacy. Teachers assessed the items on a 6-point Likert scale (1 = *totally disagree*, 6 = *totally agree*). Items were, for example, "I am snowed under with work"; "I feel dispirited at work and I think of leaving my job" and "I frequently question the value of my work."

References:

Näätänen, P., Aro, A., Matthiesen, S., & Salmela-Aro, K. (2003). *Bergen Burnout Indicator 15*. Edita.

Salmela-Aro, K., Rantanen, J., Hyvönen, K., Tilleman, K., & Feldt, T. (2011). Bergen Burnout Inventory: reliability and validity among Finnish and Estonian managers. *International Archives of Occupational and Environmental Health, 84*, 635–645. https://doi.org/10.1007/s00420-010-0594-3

Sources of Work-Related Stress and Coping Strategies with Stress

Sources of work-related stress were measured by asking them to write down their answers to an open-ended question: "What causes you the most stress and exhaustion at work?" (Elomaa et al., 2020).

Coping strategies were measured by asking them to write down their answers to an open-ended question: "What are your ways of coping with work-related stress and exhaustion?" (Aulèn et al., 2021; Elomaa et al., 2020).

References:

Aulèn, A.-M., Pakarinen, E., Feldt, T., & Lerkkanen, M.-K. (2021). Teacher coping profiles in relation to teacher well-being: A mixed-method approach. *Teaching and Teacher Education, 102*, 103323. https://doi.org/10.1016/j.tate.2021.1033230742-051X

Elomaa, M., Pakarinen, E., Eskelä-Haapanen, S., Halttunen, L., Von Suchodoletz, A. & Lerkkanen, M.-K. (2020). Directors' Stress in Day Care Centers: Related Factors and Coping Strategies. *International Journal of Educational Management*. https://doi.org/10.1108/IJEM-10-2019-0383

Teachers' Stress

Teachers' stress was measured with a question which is part of the Occupational Stress Questionnaire: "Stress means a situation in which a person feels tense, restless, nervous or anxious or is unable to sleep at night because their mind is troubled all the time. Do you feel this kind of stress these days?" (Elo et al., 2003). The single-item stress measure has been verified as acceptable for measuring variance in well-being between different groups (Aulèn et al., 2021; Elo et al., 2003; Pöysä et al., 2021). Teachers rated the item with a 6-point Likert scale (1 = *not at all*, 6 = *very much*).

The degree of teachers work related stress was further inquired with one continuation item ("To what extent do you feel that stress is being caused by your work?"). The question has been developed in TESSI study (Lerkkanen & Pakarinen, 2021) and used in a study by Aulén et al., (2022). Teachers rated the item on a 4-point Likert scale (1 = *not at all*, 4 = *entirely/fully*).

References:

Aulèn, A.-M., Pakarinen, E., Feldt, T., & Lerkkanen, M.-K. (2021). Teacher coping profiles in relation to teacher well-being: A mixed-method approach. *Teacher and Teacher Education, 102*, 103323. https://doi.org/10.1016/j.tate.2021.1033230742-051X

Aulén, A.-M., Pakarinen, E., Feldt, T., Tolvanen, A., & Lerkkanen, M.-K. (2022). Psychological detachment as a mediator between successive days' job stress and negative affect of teachers. *Frontiers in Education, 7*, 903606. https://doi.org/10.3389/feduc.2022.903606

Elo, A., Leppänen, A., & Jahkola, A. (2003). Validity of a single-item measure of stress symptoms. *Scandinavian Journal of Work, Environment & Health, 29*(6), 444–451. https://doi.org/10.5271/sjweh.752

Lerkkanen, M.-K., & Pakarinen, E. (2021). *Teacher and Student Stress and Interaction in Classroom (TESSI).* V. 31.7.2021. https://doi.org/10.17011/jyx/dataset/77741

Pöysä, S., Pakarinen, E., & Lerkkanen, M.-K. (2021). Patterns of teachers' occupational well-being during the Covid-19 pandemic: Relations to experiences of exhaustion and recovery, and interactional styles of teaching. *Frontiers in Education, 6*, 699785. https://doi.org/10.3389/feduc.2021.699785

Teachers' Self-Efficacy

Teachers' Self-Efficacy Beliefs in Language and Literacy and in Math

The teachers' self-efficacy beliefs were measured with the Teachers' Sense of Efficacy Scale (TSES; Tschannen-Moran & Woolfolk Hoy, 2001). The measure includes 24 items that are assessed on a 5-point scale (1 = *not at all*, 5 = *to a great deal*). The TSES measure includes three 8-item subscales: (1) efficacy for instructional strategies (e.g., "To what extent can you provide an alternative explanation or example when students are confused?"), (2) efficacy for classroom management (e.g., "How much can you do to control disruptive behavior in the classroom?") and (3) efficacy for student engagement (e.g., "How much can you do to motivate students who show low interest in schoolwork?"). For this study, teachers were asked to rate subscales 1 (efficacy for instructional strategies) and 3 (efficacy for student engagement) separately for math and literacy instruction so as to indicate the domain-specificity of their self-efficacy beliefs. Teachers then evaluated their efficacy for classroom management globally for their teaching in general without explicitly thinking about any specific subject.

Reference:

Tschannen-Moran, M., Woolfolk Hoy, A. & Hoy, W.K. (1998). Teacher efficacy: Its meaning and measure. *Review of Educational Research, 68*, 202–248. https://doi.org/10.3102/00346543068002202

5.7.4. TQ – School Phase: Grade 2 (T6)

TEACHERS' BACKGROUND CHARACTERISTICS

Teacher Characteristics

- Gender: (1) Female; (2) Male; (3) I prefer not to answer
- Age in years: ____
- Basic education (1) Class teacher; (2) Dual qualified class teacher; (3) Subject teacher; (4) Special education teacher; (5) Other, what: ____
- Current occupational title: (1) Class teacher; (2) Special class teacher; (3) Other, what: ____
- Have you participated in professional development over the past five years? (1) No; (2) Yes. If yes:
  - PD1 The content and duration
  - PD2 The content and duration
  - PD3 The content and duration
- Work experience: (a) In school: ____years and/or ____months; (b) Other teaching experience: ____years and/or ____months
- Work history in the current school: ____years and/or ____months
- Type of employment relationships (temporary, permanent, substitute other)

Classroom Characteristics

- Number of additional staff in the classroom: Number: ____ Professional titles: ____
- Availability of additional staff in the classroom: ____hours/week
- Number of students in the classroom: ____students, of which ____boys and ____girls
- Classroom composition: Number of children:
  - …with needs for support in some areas of learning;
  - …with needs for support in socio-emotional development;
  - …with needs for support in behavior;
  - …with needs for support in attention or concentration;
  - …who do not speak Finnish as their native language;
  - …gifted children or children who are ahead of others in development.
  - There are no aforementioned needs of support in my classroom.
- Ages of the children in the classroom: (a) 7-year-olds (born in 2014); (b) 8-year-olds (born in 2013); (c) 9-year-olds (born in 2012); (d) 10-year-olds (born in 2011); (e) Other, born in ____

TEACHERS' OCCUPATIONAL WELL-BEING

Teachers' Stress

Teachers' stress was measured with a question which is part of the Occupational Stress Questionnaire: "Stress means a situation in which a person feels tense, restless, nervous or anxious or is unable to sleep at night because their mind is troubled all the time. Do you feel this kind of stress these days?" (Elo et al., 2003). The single-item stress measure has been verified as acceptable for measuring variance in well-being between different groups (Aulèn et

al., 2021; Elo et al., 2003; Pöysä et al., 2021). Teachers rated the item with a 6-point Likert scale (1 = *not at all*, 6 = *very much*).

The degree of teachers' work-related stress was further inquired with one continuation item: "To what extent do you feel that stress is being caused by your work?" The question has been developed in TESSI study (Lerkkanen & Pakarinen, 2021) and used in a study by Aulén et al., (2022). Teachers rated the item on a 4-point Likert scale (1 = *not at all*, 4 = *entirely/fully*).

Due to the collection of data during the COVID-19 pandemic, an additional continuation item was added to understand the perceived impact of COVID-19 pandemic on teachers' stress. The degree of *teachers' COVID-19 -related stress* was inquired about with one item, adapted from Pöysä et al. (2021): "To what extent do you feel that stress is being caused by the COVID-19 pandemic?" Teachers rated the item on a 4-point Likert scale (1 = *not at all*, 4 = *entirely/ fully*).

References:

Aulèn, A.-M., Pakarinen, E., Feldt, T., & Lerkkanen, M.-K. (2021). Teacher coping profiles in relation to teacher well-being: A mixed-method approach. *Teacher and Teacher Education, 102*, 103323. https://doi.org/10.1016/j.tate.2021.1033230742-051X

Aulén, A.-M., Pakarinen, E., Feldt, T., Tolvanen, A., & Lerkkanen, M.-K. (2022). Psychological detachment as a mediator between successive days' job stress and negative affect of teachers. *Frontiers in Education, 7*, 903606. https://doi.org/10.3389/feduc.2022.903606

Elo, A., Leppänen, A., & Jahkola, A. (2003). Validity of a single-item measure of stress symptoms. *Scandinavian Journal of Work, Environment & Health, 29*(6), 444–451. https://doi.org/10.5271/sjweh.752

Lerkkanen, M.-K., & Pakarinen, E. (2021). *Teacher and Student Stress and Interaction in Classroom (TESSI).* V. 31.7.2021. https://doi.org/10.17011/jyx/dataset/77741

Pöysä, S., Pakarinen, E., & Lerkkanen, M.-K. (2021). Patterns of teachers' occupational well-being during the Covid-19 pandemic: Relations to experiences of exhaustion and recovery, and interactional styles of teaching. *Frontiers in Education, 6*, 699785. https://doi.org/10.3389/feduc.2021.699785

Work Engagement

Teachers' work engagement was measured with the Utrecht Work Engagement Scale (UWES; Schaufeli et al., 2002). The UWES consists of nine items that fall into three subscales: vigor, dedication, and absorption. These subscales capture teachers' energy at work (e.g., "At my work, I feel I am bursting with energy"), feelings of pride and enthusiasm in work (e.g., "My job inspires me"), and concentration while working (e.g., "I am immersed in my work"). Teachers rated the items with a 7-point Likert scale (1 = *never*, 7 = *daily*).

Reference:

Schaufeli, W., Salanova, M., González-Romá, V., & Bakker, A. (2002). The measurement of engagement and burnout: A two sample confirmatory factor analytic approach. *Journal of Happiness Studies, 3*, 71–92. https://doi.org/10.1023/A:1015630930326

Teachers' Self-Efficacy

Math Teaching Self-Efficacy

Teachers' math teaching self-efficacy beliefs were measured with the Mathematics Teaching Efficacy Belief Instrument (MTEBI: Enochs et al., 2000). The MTEBI is comprised of two subscales, personal mathematics teaching efficacy (PMTE, 13 items) and mathematics teaching outcome expectancy (MTOE, 8 items). For the VUOKKO study, 11 out of 13 personal mathematics teaching efficacy items were selected. Teachers rated each item on a 5-point scale (1 = *strongly agree*, 5 = *strongly disagree*). Example items: "Even if I try very hard, I will not teach mathematics as well as I will most subjects" and "I understand mathematics concepts well enough to be effective in teaching elementary mathematics."

Reference:

Enochs, L. G., Smith, P. L., & Huinker, D. (2000). Establishing factorial validity of the mathematics teaching efficacy beliefs instrument. *School Science and mathematics, 100*(4), 194–202. https://doi.org/10.1111/j.1949-8594.2000.tb17256.x

Teachers' Beliefs and Confidence about Math

Teachers reported their general beliefs and attitudes toward math with eight items, tapping both math anxiety and math enjoyment. Three items were selected from the Early Mathematics Beliefs and Confidence Survey (Chen et al., 2014) according to being appropriate for early elementary teachers in the VUOKKO study (e.g., "Just the word 'math' can make me feel nervous"; "I'm not a 'math person'" and "I like coming up with creative ways to solve math problems"). Five items were derived from the Beliefs About Mathematics and Teaching survey (MacGyvers et al., 1993; e.g., "When my answer to a math problem doesn't match someone else's, I usually assume that my answer is wrong" and "I enjoy encountering situations in my everyday life [e.g., sewing, carpentry, finances] that require me to use math to solve problems"). Teachers rated each item on a 5-point scale (1 = *strongly agree*, 5 = *strongly disagree*).

References:

Chen, J. Q., McCray, J., Adams, M., & Leow, C. (2014). A survey study of early childhood teachers' beliefs and confidence about teaching early math. Early Childhood Education Journal, 42, 367–377. https://doi.org/10.1007/s10643-013-0619-0

MacGyvers, Stipek, Salmon & Bogard (1993). *The Beliefs About Mathematics and Teaching*. Unpublished research survey. UCLA Grad. School of Education.

Teaching and Teacherhood

Teachers' Interactional Style

Teachers' interactional style in the classroom was measured using the Teachers' Interactional Style Scale (Aunola et al., 2005; Barber, 1996) used in prior studies. The scale tapped different aspects of teachers' interactional style (i.e., affection, behavioral control, and psychological control) and teaching stress. In the original scale there were 20 items, but for the current study, we used the 19-item version, formerly used in the First Steps study (Lerkkanen et al., 2006).

Example items from the Affection scale: “I often show the children of my class that I care about them” and “I often tell my pupils how much I appreciate it that they try to do something or achieve something.” From the Behavioral Control scale: “Children have to learn that rules are important in our group” and “Children should learn to behave well toward their teachers.” From the Psychological Control scale: “I believe it is good for the children in my class to know all the things I do for them” and “Children in my class need to learn to respect how good their situation is.” Teachers rated 19 items on a 5-point scale (1 = *not like me at all*, 5 = *very much like me*).

References:

Aunola, K., Lerkkanen, M.-K., Poikkeus, A.-M., & Nurmi, J.-E. (2005). *Teacher Interactional Style Scale*. University of Jyväskylä.

Barber, K. K. (1996). Parental psychological control: Revisiting a neglected construct. *Child Development, 67*, 3296–3319. https://doi.org/10.1111/j.1467-8624.1996.tb01915.x

Gerris, J. M., Vermulst, A., van Boxtel, D., Janssens, J., van Zutphen, R. & Feeling, A. (1993). *Parenting in Dutch Families*. University of Nijmegen. Institute of Family Studies.

Teaching During the COVID-19 Pandemic

Teachers responded to five items concerning the experienced impact of the COVID-19 pandemic on their teaching and classroom as a learning environment. The scale was developed in the project, with examples derived from the TESSI study (Lerkkanen & Pakarinen, 2021). Items concerned establishing successful teaching methods, communication with parents/homes, differentiation of teaching, establishing and maintaining positive classroom atmosphere, and establishing and maintaining classroom interaction that encourages learning. Teacher rated their responses on a 5-point Likert scale *(*1 = *negatively*, 3 = *not negatively nor positively*, 5 = *positively*).

Reference:

Lerkkanen, M.-K. & Pakarinen, E. (2021). *Teacher and Student Stress and Interaction in Classroom (TESSI).* V. 31.7.2021. https://doi.org/10.17011/jyx/dataset/77741

Individual Instruction in Reading and Math

Teachers were asked to rate the extent to which they provide individualized instruction to their students in the classroom. Individualized instruction was evaluated on the basis of six items (see Lerkkanen, 2003, 2006), separately for math and Finnish language and literacy: (1) Individualization by time; (2) Individualization by content depth or breadth; (3) Individualization by materials; (4) Individualization by methods/practices; (5) Individualization by organizing teaching; and (6) Individualization by assistant or special education teacher. Each item was rated on a 5-point scale (1 = *hardly ever*, 5 = *always / on every lesson*).

References:

Lerkkanen, M.-K. (2003). *Lukemaan opettamisen käytänteet 1. luokalla* [Reading instruction practices during the first grade]. Jyväskylän yliopisto. Opettajankoulutuslaitos. Julkaisematon haastattelumateriaali.

Lerkkanen, M.-K. (2006). *Lukemaan oppiminen ja opettaminen esi- ja alkuopetuksessa* [Learning to read and reading instruction in pre-primary and primary education]. WSOY.

5.7.5. TQ – School Phase: Grade 3 (T7)

TEACHERS' BACKGROUND CHARACTERISTICS

Teacher Characteristics

- Gender: (1) Female; (2) Male; (3) I prefer not to answer
- Age in years: ____
- Basic education (1) Class teacher; (2) Dual qualified class teacher; (3) Subject teacher; (4) Special education teacher; (5) Other, what: ____
- Current occupational title: (1) Class teacher; (2) Special class teacher; (3) Other, what: ____
- Have you participated in professional development over the past five years? (1) No; (2) Yes. If yes:
  - PD1 The content and duration
  - PD2 The content and duration
  - PD3 The content and duration
- Work experience: (a) In school: ____years and/or ____months; (b) Other teaching experience: ____years and/or ____months
- Work history in the current school: ____years and/or ____months
- Type of employment relationships (temporary, permanent, substitute other)

Classroom Characteristics

- Number of additional staff in the classroom: Number: ____ Professional titles: ____
- Availability of additional staff in the classroom: ____ hours/week
- Number of students in the classroom: ____students, of which ____boys and ____girls
- Classroom composition: Number of children:
  - …with needs for support in some areas of learning;
  - …with needs for support in socio-emotional development;
  - …with needs for support in behavior;
  - …with needs for support in attention or concentration;
  - …who do not speak Finnish as their native language;
  - …gifted children or children who are ahead of others in development.
  - There are no aforementioned needs of support in my classroom.
- Ages of the children in the classroom: (a) Born in 2014; (b) Born in 2013; (c) Born in 2012; (d) Born in 2011; (e) Other, born in ____

Teachers' Occupational Well-Being

Teachers' Stress

Teachers' general stress was measured with one item ("Stress means a situation in which a person feels tense, restless, nervous, or anxious, or is unable to sleep at night because their mind is troubled all the time. Do you feel this kind of stress these days?"), which is part of the Occupational Stress Questionnaire (Elo et al., 2003). Teachers rated the item with a 6-point Likert scale (1 = *not at all*, 6 = *very much*).

The degree of teachers' work-related stress was further inquired with one continuation item: "To what extent do you feel that stress is being caused by your work?" Teachers rated the item on a 4-point Likert scale (1 = *not at all*, 4 = *entirely/fully*).

Reference:

Elo, A., Leppänen, A., & Jahkola, A. (2003). Validity of a single-item measure of stress symptoms. *Scandinavian Journal of Work, Environment & Health, 29*(6), 444–451. https://doi.org/10.5271/sjweh.752

Work Engagement

Teachers' work engagement was measured with the Utrecht Work Engagement Scale (UWES; Schaufeli et al., 2002). The UWES consists of nine items that fall into three subscales: vigor, dedication, and absorption. These subscales capture teachers' energy at work (e.g., "At my work, I feel I am bursting with energy"), feelings of pride and enthusiasm in work (e.g., "My job inspires me"), and concentration while working (e.g., "I am immersed in my work"). Teachers rated the items with a 7-point Likert scale (1 = *never*, 7 = *daily*).

Reference:

Schaufeli, W., Salanova, M., González-Romá, V., & Bakker, A. (2002). The measurement of engagement and burnout: A two sample confirmatory factor analytic approach. *Journal of Happiness Studies, 3*, 71–92. https://doi.org/10.1023/A:1015630930326

Teachers' Self-Efficacy

Teachers' Self-Efficacy Beliefs in Language and Literacy and in Math

The teachers' self-efficacy beliefs were measured with the Teachers' Sense of Efficacy Scale (TSES; Tschannen-Moran & Woolfolk Hoy, 2001). The measure includes 24 items that are assessed on a 5-point scale (1 = *not at all*, 5 = *to a great deal*). The TSES measure includes three 8-item subscales: (1) efficacy for instructional strategies (e.g., To what extent can you provide an alternative explanation or example when students are confused?), (2) efficacy for classroom management (e.g., How much can you do to control disruptive behavior in the classroom?) and (3) efficacy for student engagement (e.g., How much can you do to motivate students who show low interest in schoolwork?). For this study, teachers were asked to rate subscales 1 (efficacy for instructional strategies) and 3 (efficacy for student engagement) separately for math and literacy instruction so as to indicate the domain-specificity of their self-

efficacy beliefs. Teachers then evaluated their efficacy for classroom management globally for their teaching in general without explicitly thinking about any specific subject.

Reference:

Tschannen-Moran, M., Woolfolk Hoy, A. & Hoy, W.K. (1998). Teacher efficacy: Its meaning and measure. *Review of Educational Research, 68*, 202–248. https://doi.org/10.3102/00346543068002202

Math Teaching Self-Efficacy

Teachers' math teaching self-efficacy beliefs were measured with the Mathematics Teaching Efficacy Belief Instrument (MTEBI: Enochs et al., 2000). The MTEBI is comprised of two subscales, personal mathematics teaching efficacy (PMTE, 13 items) and mathematics teaching outcome expectancy (MTOE, 8 items). For the VUOKKO study, 11 out of 13 personal mathematics teaching efficacy items were selected. Teachers rated each item on a 5-point scale (1 = *strongly agree*, 5 = *strongly disagree*). Example items: "Even if I try very hard, I will not teach mathematics as well as I will most subjects" and "I understand mathematics concepts well enough to be effective in teaching elementary mathematics."

Reference:

Enochs, L. G., Smith, P. L., & Huinker, D. (2000). Establishing factorial validity of the mathematics teaching efficacy beliefs instrument. *School Science and mathematics, 100*(4), 194–202. https://doi.org/10.1111/j.1949-8594.2000.tb17256.x

Teachers' Beliefs and Confidence about Math

Teachers reported their general beliefs and attitudes toward math with eight items, tapping both math anxiety and math enjoyment. Three items were selected from the Early Mathematics Beliefs and Confidence Survey (Chen et al., 2014) according to being appropriate for early elementary teachers in the VUOKKO study (e.g., "Just the word 'math' can make me feel nervous"; "I'm not a 'math person'" and "I like coming up with creative ways to solve math problems"). Five items were derived from the Beliefs About Mathematics and Teaching - survey (MacGyvers et al., 1993, e.g., "When my answer to a math problem doesn't match someone else's, I usually assume that my answer is wrong" and "I enjoy encountering situations in my everyday life [e.g., sewing, carpentry, finances] that require me to use math to solve problems"). Teachers rated each item on a 5-point scale (1 = *strongly agree*, 5 = *strongly disagree*).

References:

Chen, J. Q., McCray, J., Adams, M., & Leow, C. (2014). A survey study of early childhood teachers' beliefs and confidence about teaching early math. *Early Childhood Education Journal, 42*, 367–377. https://doi.org/10.1007/s10643-013-0619-0

MacGyvers, Stipek, Salmon & Bogard (1993). *The Beliefs About Mathematics and Teaching*. Unpublished research survey. UCLA Grad. School of Education.

Teaching Literacies

Teachers' perceived experiences of mastery in teaching literacies were measured with ten items. Items were derived from the Lukiloki research project and they were originally based

on the work of Aro and Björn (2016) and Peura et al. (2021). Teachers were asked to rate the extent to which they felt they were able to cover the following contents of teaching and instruction: reading development, risk factors related to reading development, reading difficulties, reading fluency, reading comprehension, critical reading, media literacy, reading of web-based texts, reading of multifaceted texts [e.g., visual or auditory texts], and using IT in teaching literacies. Teachers rated each item on a 5-point scale (1 = *not at all*, 5 = *very well*).

References:

Aro, M., & Björn, P. M. (2016). Preservice and inservice teachers' knowledge of language constructs in Finland. *Annals of Dyslexia, 66*, 111–126. https://doi.org/10.1007/s11881-015-0118-7

Peura, P., Aro, T., Räikkönen, E., Viholainen, E., Koponen, T., Usher, E. L., & Aro, M. (2021). Trajectories of change in reading self-efficacy: A longitudinal analysis of self-efficacy and its sources. *Contemporary Educational Psychology, 64*, 101947. https://doi.org/10.1016/j.cedpsych.2021.101947

TEACHING AND TEACHERHOOD

Individual Instruction in Reading and Math

Teachers were asked to rate the extent to which they provide individualized instruction to their students in the classroom. Individualized instruction was evaluated on the basis of six items (see Lerkkanen, 2003, 2006), separately for math and Finnish language and literacy: (1) Individualization by time; (2) Individualization by content depth or breadth; (3) Individualization by materials; (4) Individualization by methods/practices; (5) Individualization by organizing teaching; and (6) Individualization by assistant or special education teacher. Each item was rated on a 5-point scale (1 = *hardly ever*, 5 = *always / on every lesson*).

References:

Lerkkanen, M.-K. (2003). *Lukemaan opettamisen käytänteet 1. luokalla* [Reading instruction practices during the first grade]. Jyväskylän yliopisto. Opettajankoulutuslaitos. Julkaisematon haastattelumateriaali.

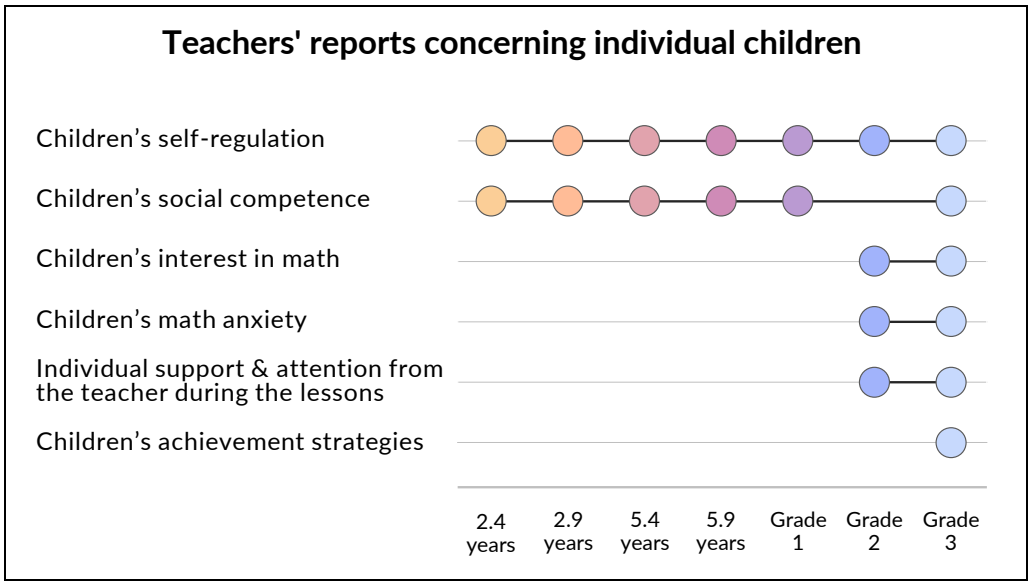
Lerkkanen, M.-K. (2006). *Lukemaan oppiminen ja opettaminen esi- ja alkuopetuksessa* [Learning to read and reading instruction in pre-primary and primary education]. WSOY.

5.8. TR – Teachers' Reports Concerning Individual Children

Teachers' reports concerning individual children were sent in T1 (2.4 years), T2 (2.9 years), T3 (5.4 years), T4 (5.9 years), T5 (Grade 1), T6 (Grade 2) and T7 (Grade 3). Figure 17 provides an overview of the measures used in each timepoint. The measures are described in detail below the figure.

T1–T2 T3–T4 T5 T6 T7

FIGURE 17 An overview of the **teachers' reports concerning individual children**.



5.8.1. TR – Toddler Phase (T1–T2)

Self-Regulation

Teachers used the Child Behavior Rating Scale (CBRS; Bronson et al., 1990) to assess children's classroom behavioral self-regulation twice during the toddler year (fall 2015 and spring 2016). The CBRS scale consists of 17 items rated on a scale from 1 (*never*) to 5 (*always*). The 17 items are spread across two subscales: The Classroom Self-Regulation subscale is comprised of 10 items that assess children's behavioral regulation during academic tasks (e.g., "Observes rules and follows directions without requiring repeated reminders" and "Concentrates when working on a task; is not easily distracted by surrounding activities") whereas the Social Skills (or Interpersonal Skills) subscale is comprised of seven items that assess children's behavioral regulation in social situations (e.g., "Willing to share toys or other things with other children when playing; does not fight or argue with playmates in disputes over property" and "Takes turns in a game situation with toys, materials, and other things without being told to do so"). In the instruction teachers were directed to think about the wording of the item "academic tasks" more broadly as "different situations in children's daily life during which the teacher guides 2- to 3-year-old children. Tasks can be related to, e.g., daily routines, meals, getting dressed or playing together with the children." Teachers were also advised to pay extra attention to two items that were presented in reverse direction. The scale was formatted as a table, with items as rows and the names of children in the classroom as columns. The 17 items were rated by the teachers on a 5-point scale (1 = *never*, 5 = *always*) for each child individually.

Classroom behavioral self-regulation has been assessed using the CBRS in prior studies (Lim et al., 2009; McClelland & Morrison, 2003).

References:

Bronson, M. B., Goodson, B. D., Layzer, J. I., & Love, J. M. (1990). *Child Behavior Rating Scale*. Abt Associates.

Lim, S. M., Rodger, S., & Brown, T. (2009). Validation of Child Behavior Rating Scale in Singapore (part 1): Rasch analysis. *Hong Kong Journal of Occupational Therapy, 20*(2), 52–62. https://doi.org/10.12968/ijtr.2009.16.5.42102

McClelland, M. M., & Morrison, F. J. (2003). The emergence of learning-related social skills in preschool children. *Early Childhood Research Quarterly, 18*, 206–224. https://doi.org/10.1016/S0885-2006(03)00026-7

Social Competence

Teachers rated toddlers' social competence twice during the toddler year (fall and spring) using the Multisource Assessment of Social Competence Scale (MASCS: Kaukiainen et al., 2005; Junttila et al., 2006). The scale consists of two dimensions, namely antisocial behavior and prosocial behavior. The dimension of prosocial behavior encompasses Cooperation (five items, e.g., "Cooperates with other children") and Empathy (three items, e.g., "Shows acceptance of other children"). The two scales falling under the dimension of antisocial behavior are Impulsivity (three items, e.g., "Has temper outbursts or tantrums") and Disruptiveness (four items, e.g., "Teases and makes fun of other children"). The original MASCS scale includes 15 items, but due to the young age of the children, two items were dropped for the Toddler Phase. These items were from the Cooperation subscale, namely "Invites other students to participate in activities" and "Is skillful in starting conversations with mates." A similar decision has been validated in a study by Panula et al. (2020) with a sample of young children. The MASCS scale was formatted as a table, with items as rows and the names of children in the classroom as columns. The 13 items were rated by the teachers on a 4-point scale (1 = *never*, 4 = *very frequently*) for each child individually.

References:

Junttila, N., Voeten, M., Kaukiainen, A., & Vauras, M. (2006). Multisource assessment of children's social competence. *Educational and Psychological Measurement, 66*, 874–895. https://doi.org/10.1177/0013164405285546

Kaukiainen, A., Junttila, N., Kinnunen, R., & Vauras, M. (2005). *MASK- Monitahoarviointi oppilaan sosiaalisesta kompetenssista* [MASCS—Multisource Assessment of Social Competence Scale, the manual]. University of Turku: Center for Learning Research & Teacher Education Department.

Panula, V., Junttila, N., Aromaa, M., Rautava, P., & Räihä, H. (2020). Parental psychosocial well-being as a predictor of the social competence of a child. *Journal of Child and Family Studies, 29*(11), 3004–3019. https://doi.org/10.1007/s10826-020-01790-6

5.8.2. TR – Preschool Phase (T3–T4)

Self-Regulation

Teachers used the Child Behavior Rating Scale (CBRS; Bronson et al., 1990) to assess children's classroom behavioral self-regulation twice during the preschool year (fall 2018 and spring 2019). The CBRS scale consists of 17 items rated on a scale from 1 (*never*) to 5 (*always*). The 17 items are spread across two subscales: The Classroom Self-Regulation subscale is comprised of 10 items that assess children's behavioral regulation during academic tasks (e.g., "Observes rules and follows directions without requiring repeated reminders" and

"Concentrates when working on a task; is not easily distracted by surrounding activities") whereas the Social Skills (or Interpersonal Skills) subscale is comprised of seven items that assess children's behavioral regulation in social situations (e.g., "Willing to share toys or other things with other children when playing; does not fight or argue with playmates in disputes over property" and "Takes turns in a game situation with toys, materials, and other things without being told to do so"). In the instruction teachers were directed to think about the wording of the item "academic tasks" more broadly as "different situations in children's daily life during which the teacher guides 5- to 6-year-old children. Tasks can be related to, e.g., daily routines, meals, getting dressed or playing together with the children." Teachers were also advised to pay extra attention to two items that were presented in reverse direction. The scale was formatted as a table, with items as rows and the names of children in the classroom as columns. The 17 items were rated by the teachers on a 5-point scale (1 = *never*, 5 = *always*) for each child individually.

Classroom behavioral self-regulation has been assessed using the CBRS in prior studies (Lim et al., 2009; McClelland & Morrison, 2003).

References:

Bronson, M. B., Goodson, B. D., Layzer, J. I., & Love, J. M. (1990). *Child Behavior Rating Scale*. Abt Associates.

Lim, S. M., Rodger, S., & Brown, T. (2009). Validation of Child Behavior Rating Scale in Singapore (part 1): Rasch analysis. *Hong Kong Journal of Occupational Therapy, 20*(2), 52–62. https://doi.org/10.12968/ijtr.2009.16.5.42102

McClelland, M. M., & Morrison, F. J. (2003). The emergence of learning-related social skills in preschool children. *Early Childhood Research Quarterly, 18*, 206–224. https://doi.org/10.1016/S0885-2006(03)00026-7

Social Competence

Teachers rated toddlers' social competence twice during the preschool year (fall and spring) using the Multisource Assessment of Social Competence Scale (MASCS: Kaukiainen et al., 2005; Junttila et al., 2006). The scale consists of two dimensions, namely antisocial behavior and prosocial behavior. The dimension of prosocial behavior encompasses Cooperation (five items, e.g., "Cooperates with other children") and Empathy (three items, e.g., "Shows acceptance of other children"). The two scales falling under the dimension of antisocial behavior are Impulsivity (three items, e.g., "Has temper outbursts or tantrums") and Disruptiveness (four items, e.g., "Teases and makes fun of other children"). The MASCS scale was formatted as a table, with items as rows and the names of children in the classroom as columns. The 15 items were rated by the teachers on a 4-point scale *(*1 = *never,* 4 = *very frequently*) for each child individually.

References:

Junttila, N., Voeten, M., Kaukiainen, A., & Vauras, M. (2006). Multisource assessment of children's social competence. *Educational and Psychological Measurement, 66*, 874–895. https://doi.org/10.1177/0013164405285546

Kaukiainen, A., Junttila, N., Kinnunen, R., & Vauras, M. (2005). *MASK- Monitahoarviointi oppilaan sosiaalisesta kompetenssista* [MASCS—Multisource Assessment of Social Competence Scale, the manual]. University of Turku: Center for Learning Research & Teacher Education Department.

5.8.3. TR – School Phase: Grade 1 (T5)

Self-Regulation

Teachers used the Child Behavior Rating Scale (CBRS; Bronson et al., 1990) to assess children's classroom behavioral self-regulation in Grade 1 spring. The 17 items are spread across two subscales: The Classroom Self-Regulation subscale is comprised of 10 items that assess children's behavioral regulation during academic tasks (e.g., "Observes rules and follows directions without requiring repeated reminders" and "Concentrates when working on a task; is not easily distracted by surrounding activities") whereas the Social Skills (or Interpersonal Skills) subscale is comprised of seven items that assess children's behavioral regulation in social situations (e.g., "Willing to share toys or other things with other children when playing; does not fight or argue with playmates in disputes over property" and "Takes turns in a game situation with toys, materials, and other things without being told to do so"). In the instruction teachers were advised to pay extra attention to two items that were presented in reverse direction. The scale was formatted as a table, with items as rows and the names of children in the classroom as columns. The 17 items were rated by the teachers on a 5-point scale (1 = *never*, 5 = *always*) for each child individually. The items were otherwise identical with prior data collection timepoints, but children were referred to as students in the instruction and in the items.

Classroom behavioral self-regulation has been assessed using the CBRS in prior studies (Lim et al., 2009; McClelland & Morrison, 2003).

References:

Bronson, M. B., Goodson, B. D., Layzer, J. I., & Love, J. M. (1990). *Child Behavior Rating Scale*. Abt Associates.

Lim, S. M., Rodger, S., & Brown, T. (2009). Validation of Child Behavior Rating Scale in Singapore (part 1): Rasch analysis. *Hong Kong Journal of Occupational Therapy, 20*(2), 52–62. https://doi.org/10.12968/ijtr.2009.16.5.42102

McClelland, M. M., & Morrison, F. J. (2003). The emergence of learning-related social skills in preschool children. *Early Childhood Research Quarterly, 18*, 206–224. https://doi.org/10.1016/S0885-2006(03)00026-7

Social Competence

Teachers rated children's social competence in Grade 1 spring using the Multisource Assessment of Social Competence Scale (MASCS: Kaukiainen et al., 2005; Junttila et al., 2006). The scale consists of two dimensions, namely antisocial behavior and prosocial behavior. The dimension of prosocial behavior encompasses Cooperation (five items, e.g., "Cooperates with other children") and Empathy (three items, e.g., "Shows acceptance of other children"). The two scales falling under the dimension of antisocial behavior are Impulsivity (three items, e.g., "Has temper outbursts or tantrums") and Disruptiveness (four items, e.g., "Teases and makes fun of other children"). The MASCS scale was formatted as a table, with items as rows and the names of children in the classroom as columns. The 15 items were rated by the teachers on a 4-point scale (1 = *never*, 4 = *very frequently*) for each child individually. The items were otherwise identical with prior data collection timepoints, but children were referred to as students in the instruction and in the items.

References:

Junttila, N., Voeten, M., Kaukiainen, A., & Vauras, M. (2006). Multisource assessment of children's social competence. *Educational and Psychological Measurement, 66*, 874–895. https://doi.org/10.1177/0013164405285546

Kaukiainen, A., Junttila, N., Kinnunen, R., & Vauras, M. (2005). *MASK- Monitahoarviointi oppilaan sosiaalisesta kompetenssista* [MASCS—Multisource Assessment of Social Competence Scale, the manual]. University of Turku: Center for Learning Research & Teacher Education Department.

5.8.4. TR – School Phase: Grade 2 (T6)

Self-Regulation

Teachers used the Child Behavior Rating Scale (CBRS; Bronson et al., 1990) to assess children's classroom behavioral self-regulation in Grade 2 spring. The 17 items are spread across two subscales: The Classroom Self-Regulation subscale is comprised of 10 items that assess children's behavioral regulation during academic tasks (e.g., "Observes rules and follows directions without requiring repeated reminders" and "Concentrates when working on a task; is not easily distracted by surrounding activities") whereas the Social Skills (or Interpersonal Skills) subscale is comprised of seven items that assess children's behavioral regulation in social situations (e.g., "Willing to share toys or other things with other children when playing; does not fight or argue with playmates in disputes over property" and "Takes turns in a game situation with toys, materials, and other things without being told to do so"). In the instruction teachers were advised to pay extra attention to two items that were presented in reverse direction. The scale was formatted as a table, with items as rows and the names of children in the classroom as columns. The 17 items were rated by the teachers on a 5-point scale (1 = *never*, 5 = *always*) for each child individually.

Classroom behavioral self-regulation has been assessed using the CBRS in prior studies (Lim et al., 2009; McClelland & Morrison, 2003).

References:

Bronson, M. B., Goodson, B. D., Layzer, J. I., & Love, J. M. (1990). *Child Behavior Rating Scale*. Abt Associates.

Lim, S. M., Rodger, S., & Brown, T. (2009). Validation of Child Behavior Rating Scale in Singapore (part 1): Rasch analysis. *Hong Kong Journal of Occupational Therapy, 20*(2), 52–62. https://doi.org/10.12968/ijtr.2009.16.5.42102

McClelland, M. M., & Morrison, F. J. (2003). The emergence of learning-related social skills in preschool children. *Early Childhood Research Quarterly, 18*, 206–224. https://doi.org/10.1016/S0885-2006(03)00026-7

Children's Interest in Math

Teachers evaluated children's individual interest in math with three items adapted from Task–Value Scale for Children (TVS-C; Nurmi & Aunola, 1999; Aunola et al., 2010). Teachers evaluated the extent to which the child likes doing math tasks, believes that they can also

perform difficult math tasks, and believes in being good at math. The teachers evaluated each child's interest on a 5-point Likert scale (1 = *never*, 5 = *always*).

References:

Aunola, K. & Nurmi, J.-E. (1999). *Task-value scale for children.* Unpublished task.

Aunola, K., Leskinen, E., & Nurmi, J.-E. (2010). Developmental dynamics between mathematical performance, task motivation, and teachers' goals during the transition to primary school. *British journal of educational psychology*, *76*(1), 21–40. https://doi.org/10.1348/000709905X51608

Children's Math Anxiety

Teachers evaluated children's individual anxiety toward math with one item. Teachers evaluated the extent to which the child feels nervous/anxious about math. Teachers evaluated each child on a 5-point Likert scale (1 = *never*, 5 = *always*).

Reference:

Developed in the project.

Individual Support and Attention from the Teacher During the Lessons

The amount of individual support from the teacher to individual children during lessons was evaluated with respect to math, reading and writing. Three items related to math (counting tasks with small numbers 1 to 20; arithmetic tasks with big numbers 20 to 1000; verbal math tasks), two items related to reading (reading fluency; reading comprehension), and two items related to writing (spelling; producing written text). The teacher was instructed to compare the amount of support or attention given to the child in question with that given to their classmates. The evaluation covered each child on each of the three skills. The teacher responded to the question: "How much individual support or attention do you give to this student during a typical lesson?" The evaluation was given on a 5-point Likert scale (1 = *clearly more than to the other students*, 5 = *clearly less than to the other students*).

Reference:

Nurmi, J.-E., Kiuru, N., Lerkkanen, M.-K., Niemi, P., Poikkeus, A.-M., Ahonen, T., Leskinen, E., & Lyyra, A. L. (2013). Teachers adapt their instruction in reading according to individual children's literacy skills. *Learning and Individual Differences, 23*, 72–79. https://doi.org/10.1016/j.lindif.2012.07.012

5.8.5. TR – School Phase: Grade 3 (T7)

Self-Regulation

Teachers used the Child Behavior Rating Scale (CBRS; Bronson et al., 1990) to assess children's classroom behavioral self-regulation in Grade 3 spring. The 17 items are spread across two subscales: The Classroom Self-Regulation subscale is comprised of 10 items that assess children's behavioral regulation during academic tasks (e.g., "Observes rules and follows directions without requiring repeated reminders" and "Concentrates when working on

a task; is not easily distracted by surrounding activities") whereas the Social Skills (or Interpersonal Skills) subscale is comprised of seven items that assess children's behavioral regulation in social situations (e.g., "Willing to share toys or other things with other children when playing; does not fight or argue with playmates in disputes over property" and "Takes turns in a game situation with toys, materials, and other things without being told to do so"). In Grade 3 spring, teachers only rated children's Classroom Self-Regulation (10 items). The scale was formatted as a table, with items as rows and the names of children in the classroom as columns. The 10 items were rated by the teachers on a 5-point scale (1 = *never*, 5 = *always*) for each child individually.

Classroom behavioral self-regulation has been assessed using the CBRS in prior studies (Lim et al., 2009; McClelland & Morrison, 2003).

References:

Bronson, M. B., Goodson, B. D., Layzer, J. I., & Love, J. M. (1990). *Child Behavior Rating Scale*. Abt Associates.

Lim, S. M., Rodger, S., & Brown, T. (2009). Validation of Child Behavior Rating Scale in Singapore (part 1): Rasch analysis. *Hong Kong Journal of Occupational Therapy, 20*(2), 52–62. https://doi.org/10.12968/ijtr.2009.16.5.42102

McClelland, M. M., & Morrison, F. J. (2003). The emergence of learning-related social skills in preschool children. *Early Childhood Research Quarterly, 18*, 206–224. https://doi.org/10.1016/S0885-2006(03)00026-7

Social Competence

Teachers rated children's social competence in Grade 3 spring using the Multisource Assessment of Social Competence Scale (MASCS: Kaukiainen et al., 2005; Junttila et al., 2006). The scale consists of two dimensions, namely antisocial behavior and prosocial behavior. The dimension of prosocial behavior encompasses Cooperation (five items, e.g., "Cooperates with other children") and Empathy (three items, e.g., "Shows acceptance of other children"). The two scales falling under the dimension of antisocial behavior are Impulsivity (three items, e.g., "Has temper outbursts or tantrums") and Disruptiveness (four items, e.g., "Teases and makes fun of other children"). The MASCS scale was formatted as a table, with items as rows and the names of children in the classroom as columns. The 15 items were rated by the teachers on a 4-point scale (1 = *never*, 4 = *very frequently*) for a child individually.

References:

Junttila, N., Voeten, M., Kaukiainen, A., & Vauras, M. (2006). Multisource assessment of children's social competence. *Educational and Psychological Measurement, 66*, 874–895. https://doi.org/10.1177/0013164405285546

Kaukiainen, A., Junttila, N., Kinnunen, R., & Vauras, M. (2005). *MASK- Monitahoarviointi oppilaan sosiaalisesta kompetenssista* [MASCS—Multisource Assessment of Social Competence Scale, the manual]. University of Turku: Center for Learning Research & Teacher Education Department.

Children's Interest in Math

Teachers evaluated children's individual interest in math with three items adapted from Task–Value Scale for Children (TVS-C; Nurmi & Aunola, 1999; Aunola et al., 2010). Teachers

evaluated the extent to which the child likes doing math tasks, believes that they can also perform difficult math tasks, and believes in being good at math. The teachers evaluated each child's interest on a 5-point Likert scale (1 = *never*, 5 = *always*).

References:

Aunola, K. & Nurmi, J.-E. (1999). *Task-value scale for children.* Unpublished task.

Aunola, K., Leskinen, E., & Nurmi, J.-E. (2010). Developmental dynamics between mathematical performance, task motivation, and teachers' goals during the transition to primary school. *British journal of educational psychology, 76*(1), 21–40. https://doi.org/10.1348/000709905X51608

Children's Math Anxiety

Teachers evaluated children's individual anxiety toward math with one item. Teachers evaluated the extent to which the child feels nervous/anxious about math. Teachers evaluated each child on a 5-point Likert scale (1 = *never*, 5 = a*lways*).

Reference:

Developed in the project.

Individual Support and Attention from the Teacher During the Lessons

The amount of individual support from the teacher to individual children during lessons was evaluated with respect to math, reading and writing. Three items related to math (arithmetic skills / basic counting skills, decimal system skills, verbal math problem solving), two items related to reading (reading fluency; reading comprehension), and two items related to writing (spelling; producing written text). The teacher was instructed to compare the amount of support or attention given to the child in question with that given to their classmates. The evaluation covered each child on each of the three skills. The teacher responded to the question: "How much individual support or attention do you give to this student during a typical lesson?" The evaluation was given on a 5-point Likert scale (1 = *clearly more than to the other students*, 5 = *clearly less than to the other students*).

Reference:

Nurmi, J.-E., Kiuru, N., Lerkkanen, M.-K., Niemi, P., Poikkeus, A.-M., Ahonen, T., Leskinen, E., & Lyyra, A. L. (2013). Teachers adapt their instruction in reading according to individual children's literacy skills. *Learning and Individual Differences, 23*, 72–79. https://doi.org/10.1016/j.lindif.2012.07.012

Achievement Strategies

Teachers were asked to evaluate individual children's achievement strategies with respect to math, reading and writing. The questions for each skill were as follows: "When difficulties appear, does the child easily start doing something else?"; "Does the child aim to persistently complete even the most challenging tasks?" and "Does the child easily give up trying?" They were adapted from the Behavior Strategy Rating Scale (BSRS; Aunola et al., 2002; Onatsu-Arvilommi & Nurmi, 2000). The teachers evaluated how likely the child behaved as described in the item separately for reading, writing and math on a 5-point Likert scale (1 = *never*, 5 = *always*).

References:

Aunola, K., Nurmi, J.-E., Niemi, P., Lerkkanen, M.-K. & Rasku-Puttonen, H. (2002). Developmental dynamics of achievement strategies, reading performance, and parental beliefs. *Reading Research Quarterly, 37 (3),* 310–327. https://doi.org/10.1598/RRQ.37.3.3

Onatsu-Arvilommi, T. & Nurmi, J.-E. (2000). The role of task-avoidant and task-focused behaviors in the development of reading and mathematical skills during the first school year: A cross-lagged longitudinal study. *Journal of Educational Psychology, 92*, 478–491. https://doi.org/10.1037/0022-0663.92.3.478

6. Coding Protocols

The tasks and observations were carefully scored and coded by the trained research assistants. The research assistants were given precise training and instructions on how to score the tasks to secure high data quality and reliability. The specific scoring principles are reported in the measure descriptions.

Certain main principles were adopted for all scoring, coding, and processing the data. These are explained below.

(1) Any unclear cases were discussed in the research group, and common decisions were made on how scoring or coding should be done. The guidelines for scoring and coding were developed through the discussions and followed systematically across the whole data.

(2) Double-checking of both scoring and coding the data into SPSS was done throughout the project. For example, of the data from Grades 1, 2 and 3 as well as of the data from the parental tasks, a randomly selected 20% of the scores were systematically double-checked. In addition, the scoring of some selected tasks (e.g., RMAT and KTLT-A) were fully double-checked due to the complexity of the scoring. The few erroneously scored or inserted values were corrected. During the data-checking process, the research group met several times to discuss and make decisions regarding the different types of scoring errors encountered. The established principles and more specific instructions for the double-checking were written down to support a more efficient process.

(3) After coding the data into the SPSS files, the distribution of the variables was checked across all data as a means to make sure that the data followed the expected distribution and that it did not include any extreme values by mistake.

(4) Missing values were coded as system missing. In addition, in some of the fluency tasks (in the skill assessment scoring), it was seen as important to separate missingness from skipping an item in the middle of a task. In these tasks the missingness due to skipping an item were marked as 99 whereas the missingness due to not reaching the end of the item list due to the given time limit were marked as system missing.

(5) Double entries in questionnaires in the cases of two adjacent numbers were coded as the mean of those numbers (e.g., responses of both 3 and 4 were coded as 3.5) in the school phase (Grades 1–3). In the ECEC phase questionnaires, both adjacent numbers were entered in the

data (e.g., responses of both 3 and 4 were coded as 34). Double entries in the cases of two not-adjacent numbers were coded as missing data (e.g., responses of both 3 and 5) in all timepoints.

(6) Voice records were gathered to support scoring of the individual skill assessments (parental assessments and Grade 3 individual assessments). The recordings were used in unclear cases and deleted after scoring.

7. Description of Variables in the Data Files

7.1. ID Creation Protocol

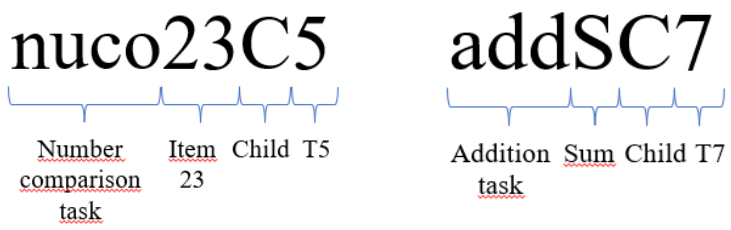
Each participant was given an individual ID number. In addition to the individual ID, a family ID was created for each participant. Family IDs link family members together. Furthermore, group/classroom numbers were created that link together the children and their teachers.

7.2. Variable Naming Protocol

A systematic protocol was used to name the variables in the data. A general principle was that the user could figure out what the specific variable represented. Another general principle was to use short variable names.

Variable Names: Tasks

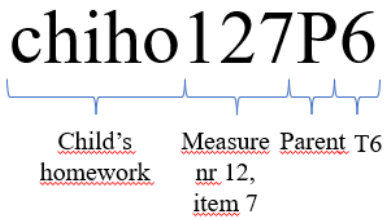
The first few (typically three or four) characters refer to the task name (e.g., nuco = number comparison; add = addition). The next one to three characters refer to either the number of the specific item of the task (e.g., 23 = the 23rd item), or in the case of the different sum scores to the type of the sum score (e.g., at = total number of attempted items, S = sum of correct responses). The next character is a capital letter that refers to the participants (e.g., C = child, P = parent), and the last number refers to the measure point (1–7). For example:



Variable Names: Questionnaires

In the variable names of the parents' and teachers' questionnaires the same principles are followed in general as in the task data. However, due to the nature of the questionnaire data, some changes were made. In the questionnaire data, the first 2–6 characters refer to the abbreviation of the questions (e.g., chiho = child's homework; chhle = child's home literacy environment). The next one to four characters varies depending on the item. Those characters refer for example to: the questionnaire numbering (e.g., 183 = the 18th measure, item number

3; 2315 = the 23rd measure, item number 15), or to the person the question was about (e.g., o = own; s = spouse) or to the content of the item (e.g., in the case of child's learning difficulties mo = delayed motor development; at = ADHD/ADD). The second-to-last character is a capital letter referring to the respondent: the letter *P* to the parent and *T* to teacher. The last character refers to the measure point (1–7). For example:



Variable Names: Observations

The variable names in the observational data follow a systematic protocol and are identical across all measurement timepoints (T2, T4, T5, T6 and T7). The variables are named according to the domains and dimensions of the Classroom Assessment Scoring System (CLASS; Pianta et al., 2008): e.g., positive_climate_respectT2; positive_climate_scoreT2; negative_climate_negative_affectT2; negative_climate_scoreT2. The variable names were constructed at the toddler phase of the VUOKKO study, while a comparable dataset was constructed for the European collaboration project (Quality Matters). To ensure the consistency across the observational CLASS data, the data variable names were maintained across the timepoints, although the overall logic does not comply with the variable naming protocol explained for the other data in the VUOKKO study.

7.3. Variable Labelling Protocol

All variables are carefully labelled in all data files. Each variable has a label that includes the variable name and a short description of the variable.

List of References:

Ahonen, T., Tuovinen, S., & Leppäsaari, T. (1999). *Nopean sarjallisen nimeämisen testi* [The test of rapid serial naming]. Niilo Mäki Instituutti & Haukkarannan koulu.

ALLEA (2017). *European code of conduct for research integrity*. https://www.allea.org/wp-content/uploads/2017/05/ALLEA-European-Code-of-Conduct-for-Research-Integrity-2017.pdf

Aro, M., & Björn, P. M. (2016). Preservice and inservice teachers' knowledge of language constructs in Finland. *Annals of Dyslexia, 66*, 111–126. https://doi.org/10.1007/s11881-015-0118-7

Ashcraft, M. H. (2002). Math Anxiety: Personal, Educational, and Cognitive Consequences. Current Directions in Psychological Science, 11(5), 181–185. https://doi.org/10.1111/1467-8721.00196

Aulén, A.-M., Pakarinen, E., Feldt, T., Tolvanen, A., & Lerkkanen, M.-K. (2022). Psychological detachment as a mediator between successive days' job stress and

negative affect of teachers. *Frontiers in Education, 7*, 903606. https://doi.org/10.3389/feduc.2022.903606

Aulèn, A.-M., Pakarinen, E., Feldt, T., & Lerkkanen, M.-K. (2021). Teacher coping profiles in relation to teacher well-being: A mixed-method approach. *Teacher and Teacher Education, 102*, 103323. https://doi.org/10.1016/j.tate.2021.1033230742-051X

Aunola, K. & Nurmi, J.-E. (1999). *Task-value scale for children.* Unpublished task.

Aunola, K., & Räsänen, P. (2007). *The Basic Arithmetic Test*. Unpublished task.

Aunola, K., Lerkkanen, M.-K., Poikkeus, A.-M., & Nurmi, J.-E. (2005). *Teacher Interactional Style Scale*. University of Jyväskylä.

Aunola, K., Leskinen, E., & Nurmi, J.-E. (2010). Developmental dynamics between mathematical performance, task motivation, and teachers' goals during the transition to primary school. *British journal of educational psychology*, *76*(1), 21–40. https://doi.org/10.1348/000709905X51608

Aunola, K., Nurmi, J.-E., Onatsu-Arvilommi, T., & Pulkkinen, L. (1999). The role of parents' self-esteem, mastery-orientation and social background in their parenting styles. *Scandinavian Journal of Psychology, 40*(4), 307–317. https://doi.org/10.1111/1467-9450.404131

Aunola, K., Nurmi, J.-E., Niemi, P., Lerkkanen, M.-K. & Rasku-Puttonen, H. (2002). Developmental dynamics of achievement strategies, reading performance, and parental beliefs. *Reading Research Quarterly, 37 (3),* 310–327. https://doi.org/10.1598/RRQ.37.3.3

Barber, K. K. (1996). Parental psychological control: Revisiting a neglected construct. *Child Development, 67*, 3296–3319. https://doi.org/10.1111/j.1467-8624.1996.tb01915.x

Blevins-Knabe, B., & Musun-Miller, L. (1996). Number use at home by children and their parents and its relationship to early mathematical performance. *Early Development and Parenting: An International Journal of Research and Practice, 5*(1), 35–45. https://doi.org/10.1002/(SICI)1099-0917(199603)5:1<35::AID-EDP113>3.0.CO;2-0

Bloom, P. J. (2010). *Measuring work attitudes in the early childhood setting. Technical manual for the Early Childhood Job Satisfaction Survey and Early Childhood Work Environment Survey* (2nd ed.). McCormick Center for Early Childhood Leadership.

Brankaer, C., Ghesquière, P., & De Smedt, B. (2017). Symbolic magnitude processing in elementary school children: A group administered paper-and-pencil measure (SYMP Test). *Behavior Research Methods, 49*, 1361–1373. https://doi.org/10.3758/s13428-016-0792-3

Bronson, M. B., Goodson, B. D., Layzer, J. I., & Love, J. M. (1990). *Child Behavior Rating Scale*. Abt Associates.

Cadima, J., Aguiar, C., Guedes, C., Wysłowska, O., Salminen, J., Slot, P., Barata, C., & Lerkkanen, M.-K. (2022). Process Quality in Toddler Classrooms in Four European Countries. *Early Education and Development, 34*(7), 1565–1589. https://doi.org/10.1080/10409289.2022.2139548

Cadima, J., Guedes, C., Aguiar, T., Slot, P., Wyslowska, O., Salminen, J., & Lerkkanen, M.-K. (2018). *Activity setting measure*. Unpublished manual. Portuguese Foundation for Science and Technology (grant PTDC/MHC-CED/5913/2014, research project "Quality Matters" University of Porto, Portugal).

Carless, S. A., Wearing, A. J., & Mann, L. (2000). A short measure of transformational leadership. *Journal of Business and Psychology, 14*, 389–405. https://doi.org/10.1023/A:1022991115523

Chen, J. Q., McCray, J., Adams, M., & Leow, C. (2014). A survey study of early childhood 1teachers' beliefs and confidence about teaching early math. *Early Childhood Education Journal, 42*, 367–377. https://doi.org/10.1007/s10643-013-0619-0

Corsi, P.M. (1972). *Human memory and the medial temporal region of the brain* [Doctoral Dissertation, McGill University]. Retrieved from https://escholarship.mcgill.ca/concern/theses/05741s554

Data Protection Act (1050/2018). https://www.finlex.fi/en/laki/kaannokset/2018/en20181050

Denckla, M. B., & Rudel, R. G. (1976). Rapid 'automatized' naming (RAN): Dyslexia differentiated from other learning disabilities. *Neuropsychologia, 14*(4), 471–479.

Dunn, L., & Dunn, L. (1981). *The Peabody picture vocabulary test-revised*. American Guidance Service.

Edwards A. A., Daucourt M. C., Hart S. A., Schatschneider C. (2022). Measuring reading anxiety in college students. *Reading and Writing, 36*(5), 1145–1180. https://doi.org/10.1007/s11145-022-10324-z

Elo, A., Leppänen, A., & Jahkola, A. (2003). Validity of a single-item measure of stress symptoms. *Scandinavian Journal of Work, Environment & Health, 29*(6), 444–451. https://doi.org/10.5271/sjweh.752

Elomaa, M., Pakarinen, E., Eskelä-Haapanen, S., Halttunen, L., Von Suchodoletz, A. & Lerkkanen, M.-K. (2020). Directors' Stress in Day Care Centers: Related Factors and Coping Strategies. *International Journal of Educational Management*. https://doi.org/10.1108/IJEM-10-2019-0383

Elomäki, T., Huolila, R., Poskiparta, E., & Saranpää, P. (1999). *Kouluvalmiuden arviointi ryhmässä. Ryhmätutkimuksen käsikirja ja seurantatutkimus* [Group assessment of school readiness. Handbook of group assessment and a follow-up study] (Publications of the Social and Health Center, City of Turku, Finland: A, 2). Social and Health Center.

Enochs, L. G., Smith, P. L., & Huinker, D. (2000). Establishing factorial validity of the mathematics teaching efficacy beliefs instrument. *School Science and mathematics, 100*(4), 194–202. https://doi.org/10.1111/j.1949-8594.2000.tb17256.x

Falanga, K., Gonida, E., & Stamovlasis, D. (2023). Predicting different types of parental involvement in children's homework: the role of parent motivational beliefs and parent affect. *European Journal of Psychology of Education, 38*(1), 249–268. https://doi.org/10.1007/s10212-022-00613-0

Finnish National Board on Research Integrity (2012). *Responsible conduct of research and procedures for handling allegations of misconduct in Finland* (Guidelines of the Finnish Advisory Board on Research Integrity 2012). Finnish National Board on Research Integrity. https://tenk.fi/sites/tenk.fi/files/HTK_ohje_2012.pdf

Francis, D., Nation, K., & McArthur, G. (2020). *The Macquarie Oxford Reading Anxiety Test-Parents (MoRAT-P)*. Motif.

General Data Protection Regulation (2016). https://eur-lex.europa.eu/eli/reg/2016/679/oj

Gerris, J. M., Vermulst, A., van Boxtel, D., Janssens, J., van Zutphen, R., & Felling, A. (1993). *Parenting in Dutch families*. University of Nijmegen. Institute of Family Studies.

Gonida, E. N., & Cortina, K. (2014). Parent involvement in homework: Relations with parent and student achievement-related motivational beliefs and achievement. *British Journal of Educational Psychology*, *84*, 376–396. https://doi.org/10.1111/bjep.12039

Hannula, M. M., & Lehtinen, E. (2005). Spontaneous focusing on numerosity and mathematical skills of young children. *Learning and Instruction*, *15*(3), 237–256. https://doi.org/10.1016/j.learninstruc.2005.04.005

Harms, T., Clifford, R., & Cryer, D. (1998). *Early childhood environment rating scale–Revised*. Teachers College Press

Harms, T., Cryer, D., & Clifford, R.M. (2003). *Infant/Toddler Environment Rating Scale–Revised*. Teachers College Press.

Hart, S. A., Ganley, C. M., & Purpura, D. J. (2016). Understanding the home math environment and its role in predicting parent report of children's math skills. *PloS One*, *11*(12), e0168227. https://doi.org/10.1371/journal.pone.0168227

Heikkilä, R., Huotari, S., Salmi, P., Korpivaara, P., Karhunen, R., Torppa, M., Aro, M., & Ahonen, T. *Nopean nimeämisen testi – RAN* [Test of rapid automized naming – RAN] (measure in preparation). Niilo Mäki Instituutti.

Hoover-Dempsey, K. V., & Sandler, H. M. (1995). Parental involvement in children's education: Why does it make a difference? *Teachers College Record*, *97*(2), 310–331. https://doi.org/10.1177/016146819509700202

Junttila, N., Voeten, M., Kaukiainen, A., & Vauras, M. (2006). Multisource assessment of children's social competence. *Educational and Psychological Measurement*, *66*, 874–895. https://doi.org/10.1177/0013164405285546

Kail, R. V., & Ferrer, E. (2007). Processing speed in childhood and adolescence: Longitudinal models for examining developmental change. *Child development*, *78*(6), 1760–1770. https://doi.org/10.1111/j.1467-8624.2007.01088.x

Kaukiainen, A., Junttila, N., Kinnunen, R., & Vauras, M. (2005). *MASK- Monitahoarviointi oppilaan sosiaalisesta kompetenssista* [MASCS—Multisource Assessment of Social Competence Scale, the manual]. University of Turku: Center for Learning Research & Teacher Education Department.

Kessels, R. P., van Den Berg, E., Ruis, C., & Brands, A. M. (2008). The backward span of the Corsi Block-Tapping Task and its association with the WAIS-III Digit Span. *Assessment*, *15*(4), 426–434.

Kessels, R.P.C., van Zandvoort, M.J.E., Postman, A., Kapelle, L.J., & de Hand, E.H.F. (2000). The Corsi Block-Tapping Task: Standardization and Normative Data. *Applied Neuropsychology, 7(4)*, 252–258.

Koponen, 2020. Unpublished questionnaire.

Koponen, T. (2018). *Number relations task.* Unpublished task

Koponen, T. (2018a). *Writing numbers*. Unpublished task.

Koponen, T. (2018b). *Number comparison: multidigit numbers*. Unpublished task.

Koponen, T. (2020). *Motor speed-copying digits.* Unpublished task.

Koponen, T. (2021a). *Number relations task*. Unpublished task.

Koponen, T. (2021b). *Number relation fluency test*. Unpublished task.

Koponen, T., & Aro, M. (2016). *Word span task.* Unpublished test material.

Koponen, T., & Mononen, R. (2010a). *The 2-minute addition fluency test*. Unpublished task.

Koponen, T., & Mononen, R. (2010b). *The 2-minute subtraction fluency test*. Unpublished task.

Koponen, T., & Mononen, R. (2010c). *The 2-minute multiplication fluency tes*t. Unpublished task.

Koponen, T., & Salminen, J. (2016). *Verbal counting test*. Unpublished task.

Koponen, T., Salminen, J., Aunio, P., & Polet, J. (2011). *LukiMat - Oppimisen arviointi: Matematiikan tuen tarpeen tunnistamisen välineet esikouluun. Käyttäjän opas* [LukiMat – Evaluation of learning: Tools for identifying the need of support in math on pre-primary education. User manual]. Niilo Mäki Institute. http://www.lukimat.fi/lukimat-oppimisen-arviointi/materiaalit/tuen-tarpeen-tunnistaminen/esiopetus/matematiikka/kayttajan-opas

Korkman, M., Kirk, U., & Kemp, S. L. (1998). *NEPSY. A developmental neuropsychological assessment.* Psychological Corporation.

La Paro, K. M., Hamre, B. K., & Pianta, R. C. (2012). *Classroom assessment scoring system. Manual, Toddler.* Paul H. Brookes Publishing Company.

LeFevre, J. A., Skwarchuk, S. L., Smith-Chant, B. L., Fast, L., Kamawar, D., & Bisanz, J. (2009). Home numeracy experiences and children's math performance in the early school years. *Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement, 41*(2), 55. https://doi.org/10.1037/a0014532

Lefly, D. L., & Pennington, B. F. (2000). Reliability and validity of the adult reading history questionnaire. *Journal of Learning Disabilities*, *33*(3), 286–296. https://doi.org/10.1177/002221940003300306

Leinonen, S., Müller, K., Leppänen, P. H., Aro, M., Ahonen, T., & Lyytinen, H. (2001). Heterogeneity in adult dyslexic readers: Relating processing skills to the speed and accuracy of oral text reading. *Reading and Writing, 14*(3–4), 265–296. https://doi.org/10.1023/A:1011117620895

Lerkkanen, M.-K. (2003). *Lukemaan opettamisen käytänteet 1. luokalla* [Reading instruction practices during the first grade]. Jyväskylän yliopisto. Opettajankoulutuslaitos. Julkaisematon haastattelumateriaali.

Lerkkanen, M.-K. (2006). *Lukemaan oppiminen ja opettaminen esi- ja alkuopetuksessa* [Learning to read and reading instruction in pre-primary and primary education]. WSOY.

Lerkkanen, M.-K., & Pakarinen, E. (2021). *Teacher and Student Stress and Interaction in Classroom (TESSI).* V. 31.7.2021. https://doi.org/10.17011/jyx/dataset/77741

Lerkkanen, M.-K., Niemi, P., Poikkeus, A.-M., Poskiparta, E., Siekkinen, M., & Nurmi, J.-E. (2006–2016). *The First Steps Study* [Alkuportaat]. University of Jyväskylä.

Lerkkanen, M.-K., Poikkeus, A.-M., & Ketonen, R. (2006). *ARMI—Luku- ja kirjoitustaidon arviointimateriaali 1. luokalle* [ARMI—A tool for assessing reading and writing skills in grade 1]. WSOY.

Lim, S. M., Rodger, S., & Brown, T. (2009). Validation of Child Behavior Rating Scale in Singapore (part 1): Rasch analysis. *Hong Kong Journal of Occupational Therapy, 20*(2), 52–62. https://doi.org/10.12968/ijtr.2009.16.5.42102

Lindeman, J. (1998). *ALLU—Ala-asteen lukutesti* [Reading test for primary school]. University of Turku.

Lohvansuu, K., Torppa, M., Ahonen, T., Eklund, K., Hämäläinen, J. A., Leppänen, P. H. T., & Lyytinen, H. (2021). Unveiling the mysteries of dyslexia—Lessons learned from the prospective Jyväskylä longitudinal study of dyslexia. *Brain Science, 11, 427.* https://doi.org/10.3390/brainsci11040427

Lonigan, C. J., Wagner, R. K., Torgesen, J. K., & Rashotte, C. A. (2007). *TOPEL: Test of preschool early literacy*. ProEd.

MacGyvers, Stipek, Salmon & Bogard (1993). *The Beliefs About Mathematics and Teaching*. Unpublished research survey. UCLA Grad. School of Education.

Maio, R., Guichard, S., & Cadima, J. (2022). In what conditions are intercultural practices implemented in disadvantaged and diverse settings in Portugal? Associations with professional and organization-related variables. *Social Psychology of Education, 25*(2–3), 509–534. https://doi.org/10.1007/s11218-022-09687-6

Martin, R. P. (2014). *Survey of children's individual differences* (SIDCA, Early Childhood Version). Unpublished Manual.

McClelland, M. M., & Morrison, F. J. (2003). The emergence of learning-related social skills in preschool children. *Early Childhood Research Quarterly, 18*, 206–224. https://doi.org/10.1016/S0885-2006(03)00026-7

McClelland, M. M., Cameron, C. E., Duncan, R., Bowles, R. P., Acock, A. C., Miao, A., & Pratt, M. E. (2014). Predictors of early growth in academic achievement: The head-toes-knees-shoulders task. *Frontiers in psychology, 5*, 81720. https://doi.org/10.3389/fpsyg.2014.00599

Midgley, C., Maehr, M. L., Hruda, L., Anderman, E. M., Anderman, L., Freeman, K. E., Gheen, M., Kaplan, A., Kumar, R., Middleton, M. J., Nelson, J., Roeser, R., & Urdan, T. (2000). *Manual for the Patterns of Adaptive Learning Scales (PALS).* University of Michigan.

Mulder, H., Hoofs, H., Verhagen, J., van der Veen, I., & Leseman, P. P. (2014). Psychometric properties and convergent and predictive validity of an executive function test battery for two-year-olds. *Frontiers in Psychology, 5*, 733. https://doi.org/10.3389/fpsyg.2014.00733

Nevala, J., Kairaluoma, L., Ahonen, T., Aro, M., & Holopainen, L. (2006). *Lukemis- ja kirjoittamistaitojen yksilötestistö nuorille ja aikuisille* [Individual reading and writing test battery for young people and adults]. Niilo Mäki Institute.

Nguyen, T. Q., Martinez-Lincoln, A., & Cutting, L. E. (2022). Tracking familial history of reading and math difficulties in children's academic outcomes. *Frontiers in Psychology, 12*, 710380. https://doi.org/10.3389/fpsyg.2021.710380

Nicholls, J. G. (1978). The development of the concepts of effort and ability, perception of academic attainment and the understanding that difficult tasks require more ability. *Child Development, 49*, 800–814. https://doi.org/10.2307/1128250

Nosworthy, N., Bugden, S., Archibald, L., Evans, B., & Ansari, D. (2013). A two-minute paper-and-pencil test of symbolic and nonsymbolic numerical magnitude processing explains variability in primary school children's arithmetic competence. *PLoS ONE 8*(7): e67918. https://doi.org/10.1371/journal.pone.0067918

Núñez-Peña, M. I., Guilera, G., & Suárez-Pellicioni, M. (2014). The Single-Item Math Anxiety Scale: An Alternative Way of Measuring Mathematical Anxiety. Journal of Psychoeducational Assessment, 32(4), 306–317. https://doi.org/10.1177/0734282913508528

Nurmi, J.-E., Kiuru, N., Lerkkanen, M.-K., Niemi, P., Poikkeus, A.-M., Ahonen, T., Leskinen, E., & Lyyra, A. L. (2013). Teachers adapt their instruction in reading according to individual children's literacy skills. *Learning and Individual Differences, 23*, 72–79. https://doi.org/10.1016/j.lindif.2012.07.012

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.* https://doi.org/10.1016/j.learninstruc.2005.04.009

Nurmi, J.-E., & Aunola, K. (1998). *Observer Rating Scale of Achievement Strategies (OSAS).* Unpublished measurement instrument.

Näätänen, P., Aro, A., Matthiesen, S., & Salmela-Aro, K. (2003). *Bergen Burnout Indicator 15*. Edita.

O'Sullivan, R. H., Chen, Y.-C., & Fish, M. C. (2014). Parental mathematics homework involvement of low-income families [TM1] with middle school students. *School Community Journal, 24*(2), 165–188. https://psycnet.apa.org/record/2014-56242-008

Onatsu-Arvilommi, T. & Nurmi, J.-E. (2000). The role of task-avoidant and task-focused behaviors in the development of reading and mathematical skills during the first school year: A cross-lagged longitudinal study. *Journal of Educational Psychology, 92*, 478–491. https://doi.org/10.1037/0022-0663.92.3.478

Pakarinen, E., Lerkkanen, M.-K., Poikkeus, A.-M., Kiuru, N., Siekkinen, M., Rasku-Puttonen, H., & Nurmi, J.-E. (2010). A validation of the classroom assessment scoring system in Finnish kindergartens. *Early Education & Development, 21*(1), 95–124. https://doi.org/10.1080/10409280902858764

Panula, V., Junttila, N., Aromaa, M., Rautava, P., & Räihä, H. (2020). Parental psychosocial well-being as a predictor of the social competence of a child. *Journal of Child and Family Studies, 29*(11), 3004–3019. https://doi.org/10.1007/s10826-020-01790-6

Pekrun, R., Goetz, T., Titz, W., & Perry, R. P. (2002). Academic emotions in students' self-regulated learning and achievement: A program of qualitative and quantitative research. *Educational Psychologist, 37,* 91–105. https://doi.org/10.1207/S15326985EP3702_4

Peura, P., Aro, T., Räikkönen, E., Viholainen, E., Koponen, T., Usher, E. L., & Aro, M. (2021). Trajectories of change in reading self-efficacy: A longitudinal analysis of self-efficacy and its sources. *Contemporary Educational Psychology, 64,* 101947. https://doi.org/10.1016/j.cedpsych.2021.101947

Pianta, R. C., K. M. La Paro, & Hamre, B. K. (2008). *Classroom Assessment Scoring System. Manual, Pre-K.* Paul H. Brookes Publishing Company.

Pianta, R. C., K. M. La Paro, & Hamre, B. K. (2008). *Classroom Assessment Scoring System. Manual, K-3.* Paul H. Brookes Publishing Company.

Poskiparta, E., Niemi, P., & Lepola, J. (1994). *Diagnostiset testit 1. Lukeminen ja kirjoittaminen* [Diagnostic tests 1. Reading and writing]. University of Turku, Centre for Research on Learning.

Pre-COOL Cohortonderzoek (2012). *Technisch rapport tweejarigen onderzoek, eerste meting 2010–2011* [Pre-COOL Cohort Study. Technical report two-year-olds' cohort, first measurement wave 2010–2011]. Kohnstamm Instituut.

Pöysä, S., Pakarinen, E., & Lerkkanen, M.-K. (2021). Patterns of teachers' occupational well-being during the Covid-19 pandemic: Relations to experiences of exhaustion and

recovery, and interactional styles of teaching. *Frontiers in Education, 6,* 699785. https://doi.org/10.3389/feduc.2021.699785.

Repo, L., Paananen, M., Eskelinen, M., Mattila, V., Lerkkanen, M-K., Gammelgård, L., Ulvinen, J., Marjanen, J., Kivistö, A., & Hjelt, H. (2019). *Varhaiskasvatuksen laatu arjessa. Varhaiskasvatussuunnitelmien toteutuminen päiväkodeissa ja perhepäivähoidossa* [Every-day quality in early childhood education and care – ECEC curriculum implementation at day-care centers and in family day-care] (Julkaisut 15:2019). Kansallinen koulutuksen arviointikeskus [Finnish National Education Evaluation Centre].

Ribner, A. D., & McHarg, G. (2021). Screens across the pond: findings from longitudinal screen time research in the US and UK. *Infant Behavior and Development, 63*, 101551. https://doi.org/10.1016/j.infbeh.2021.101551

Räsänen, P., & Leino, L. (2005). *KTLT-laskutaidon testi* [KTLT-a test of basic mathematical skills]. Niilo Mäki instituutti.

Räsänen, P., Linnanmäki, K., Haapamäki, C., & Skagersten, D. (2008). *RMAT-Test av räknefärdighet hos evelver i åldern 9–12 år* [A mathematical achievement test for ages 9–12 in Finnish-Swedish]. Niilo Mäki Institute.

Salmela-Aro, K., Rantanen, J., Hyvönen, K., Tilleman, K., & Feldt, T. (2011). Bergen Burnout Inventory: reliability and validity among Finnish and Estonian managers. *International Archives of Occupational and Environmental Health, 84*, 635–645. https://doi.org/10.1007/s00420-010-0594-3

Schaufeli, W., Salanova, M., González-Romá, V., & Bakker, A. (2002). The measurement of engagement and burnout: A two sample confirmatory factor analytic approach. *Journal of Happiness Studies, 3*, 71–92. https://doi.org/10.1023/A:1015630930326

Schrank, F. A. (2006). Specification of the cognitive processes involved in performance on the Woodcock-Johnson III. Assessment Service Bulletin, (7). Riverside Publishing.

Sen, T., & Spruyt, K. (2020). Pediatric sleep tools: An updated literature review. *Frontiers in Psychiatry, 11*, 317. https://doi.org/10.3389/fpsyt.2020.00317

Silinskas, G., Kiuru, N., Aunola, K., Lerkkanen, M.-K., & Nurmi, J.-E. (2015). The developmental dynamics of children's academic performance and mothers' homework-related affect and practices. *Developmental Psychology*, *51*(4), 419. https://doi.org/10.1037/a0038908

Slot, P., Cadima, J., Salminen, J., Pastori, G., & Lerkkanen, M.-K. (2016). *Multiple case study in seven European countries regarding culture-sensitive classroom quality assessment.* WP2.3 Curriculum and quality analysis impact review. CARE project. Utrecht University. https://ecec-care.org/fileadmin/careproject/Publications/reports/CARE_WP2_D2_3_Multiple_Case_study_FINAL_REPORT.pdf

Slot, P.L., Leseman, P.P.M., Mulder, H., & Verhagen, J. (2015). Associations between structural quality aspects and process quality in Dutch early childhood education and care settings. *Early Childhood Research Quarterly, 33*, 64–76. https://doi.org/10.1016/j.ecresq.2015.06.001

Smith-Donald, R., Raver, C. C., Hayes, T., & Richardson, B. (2007). Preliminary construct and concurrent validity of the pre-school self-regulation assessment (PSRA) for field-based research. *Early Childhood Research Quarterly, 22(2),* 173–187. https://doi.org/10.1016/j.ecresq.2007.01.002

Soininen, V., Pakarinen, E., & Lerkkanen, M.-K. (2023). Reciprocal associations among teacher–child interactions, teachers' work engagement, and children's social competence. *Journal of Applied Developmental Psychology, 85*, 101508. https://doi.org/10.1016/j.appdev.2022.101508

Sorvo, R., Koponen, T., Viholainen, H., Aro, T., Räikkönen, E., Peura, P., Dowker, A., & Aro, M. (2017). Math anxiety and its relationship with basic arithmetic skills among primary school children. *British Journal of Educational Psychology, 87*(3), 309–327. https://doi.org/10.1111/bjep.12151

Sorvo, R., Koponen, T., Viholainen, H., Aro, T., Räikkönen, E., Peura, P., Tolvanen, A., & Aro, M. (2019). Development of math anxiety and its longitudinal relationships with arithmetic achievement among primary school children. *Learning and Individual Differences, 69*, 173–181. https://doi.org/10.1016/j.lindif.2018.12.005

Stefansson, H., Meyer-Lindenberg, A., Steinberg, S., Magnusdottir, B., Morgen, K., Arnarsdottir, S., Bjornsdottir, G., Bragi Walters, G., Jonsdottir, G. A., Doyle, O. M., Tost, H., Grimm, O., Kristjansdottir, S., Snorrason, H., Davidsdottir, S. R., Gudmundsson, L. J., Jonsson, G. F, Stefansdottir, B., Helgadottir, I., Haraldsson, M., Jonsdottir, B.,... & Tost, H. (2014). CNVs conferring risk of autism or schizophrenia affect cognition in controls. *Nature*, *505*(7483), 361. https://doi.org/10.1038/nature12818

Sury, D., & Gaab, N. (2020). The Adult Arithmetic History Questionnaire. https://psyarxiv.com/zt6ku/

Torppa, M., Poikkeus, A.-M., Laakso, M. L., Eklund, K., & Lyytinen, H. (2006). Predicting delayed letter knowledge development and its relation to grade 1 reading achievement among children with and without familial risk for dyslexia. *Developmental Psychology, 42*(6), 1128–1142. *https*://doi.org/10.1037/0012-1649.42.6.1128

Tschannen-Moran, M. & Woolfolk Hoy, A. (2001). Teacher efficacy: capturing an elusive construct. *Teaching and Teacher Education 17*, 783–805. https://doi.org/10.1016/S0742-051X(01)00036-1

Tschannen-Moran, M., Woolfolk Hoy, A. & Hoy, W.K. (1998). Teacher efficacy: Its meaning and measure. *Review of Educational Research, 68*, 202–248. https://doi.org/10.3102/00346543068002202

Tunkkari, M., Aunola, K., Hirvonen, R., Silinskas, G., & Kiuru, N. (2021). The interplay between maternal homework involvement, task-avoidance, and achievement among adolescents. *Journal of Family Psychology*, *35*(7), 863. https://doi.org/10.1037/fam0000686

Viikki, M., & Leinonen, E. (2015). Ahdistuneisuushäiriöiden diagnostiikka ja lääkehoito perusterveydenhuollossa. *Aikakauskirja Duodecim, 131(6),* 583–590 https://www.duodecimlehti.fi/lehti/2015/6/duo12168

Wagner, R. K., Torgesen, J. K., Rashotte, C. A., & Pearson, N. A. (2009). *TOSREC: Test of sentence reading efficiency and comprehension.* Pro-Ed.

Walker, J. M., Wilkins, A. S., Dallaire, J. R., Sandler, H. M., & Hoover-Dempsey, K. V. (2005). Parental involvement: Model revision through scale development. *The Elementary School Journal*, *106*(2), 85–104. https://doi.org/10.1086/499193

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS Scales. *Journal of Personality & Social Psychology, 54*, 1063–1070. https://doi.org/10.1037/0022-3514.54.6.1063

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., Bonino da Silva, L., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosac, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R.,…& Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific data*, *3*(1), 1–9. https://doi.org/10.1038/sdata.2016.18

WMA (2013). *WMA Declaration of Helsinki* (1964/2013). Ethical principles for medical research involving human subjects. http://www.wma.net/en/30publications/10policies/b3/

Woodcock, R. W., & Johnson, M. B. (1977). *Woodcock-Johnson Psycho-Educational Battery*. Riverside Publishing.

Wright, R. J., Martland, J., & Stafford, A. K. (2006). *Early numeracy: Assessment for teaching and intervention.* Sage.

Wynn, K. (1990). Children's understanding of counting. *Cognition*, *36*, 155–193. https://doi.org/10.1016/0010-0277(90)90003-3

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

Zhang, X., Nurmi, J.-E., Kiuru, N., Lerkkanen, M.-K., & Aunola, K. (2011). A teacher-report measure of children's task-avoidant behavior: A validation study of the behavioral strategy rating scale. *Learning and Individual Differences*, *2*, 690–698. https://doi.org/10.1016/j.lindif.2011.09.007