Quality in Icelandic mathematics teaching

Cognitive activation in mathematics lessons
in a Nordic context

Jóhann Örn Sigurjónsson

Thesis for the degree of Philosophiae Doctor

January 2023

School of Education

UNIVERSITY OF ICELAND

Quality in Icelandic mathematics teaching

Cognitive activation in mathematics lessons in a Nordic context

Jóhann Örn Sigurjónsson

Thesis for the degree of Philosophiae Doctor

Supervisors

Dr. Anna Kristín Sigurðardóttir
Dr. Berglind Gísladóttir

Doctoral committee

Dr. Jorryt van Bommel

Opponents at defence

Dr. Eckhard Klieme
Dr. Kjersti Wæge

January 2023

School of Education

UNIVERSITY OF ICELAND

Gæði kennslu í stærðfræði á Íslandi

Hugræn virkjun í kennslustundum í stærðfræði í norrænu samhengi

Jóhann Örn Sigurjónsson

Ritgerð til doktorsgráðu

Leiðbeinendur

Dr. Anna Kristín Sigurðardóttir
Dr. Berglind Gísladóttir

Doktorsnefnd

Dr. Jorryt van Bommel

Andmælendur

Dr. Eckhard Klieme
Dr. Kjersti Wæge

Janúar 2023

Menntavísindasvið

HÁSKÓLI ÍSLANDS

ISBN 978-9935-9663-0-8

ORCID: 0000-0003-1069-6227

Reykjavik, Iceland 2023

Ágrip

Í rannsókninni sem þessi doktorsritgerð byggir á var miðað að því að þróa skilning á hugrænni virkjun í stærðfræðikennslu á Íslandi og á Norðurlöndum með því að rýna kerfisbundið í myndbandsupptökur úr kennslustundum og greina nemendakönnun. Tíu stærðfræðikennarar á Íslandi tóku þátt. Þrjár til fjórar aðliggjandi kennslustundir í 8. bekk voru teknar upp á myndband. Í heild voru 34 kennslustundir greindar með greiningarrammanum PLATO. Stærðfræðileg viðfangsefni voru greind með The Task Analysis Guide. Nemendur á aldrinum 13-14 ára (N = 217) svöruðu Tripod nemendakönnuninni og svör þeirra voru borin saman við athugunargildi út frá myndbandsgreiningu með PLATO. Til nánari greiningar voru valdar kennslustundir þar sem hugræn virkjun var metin á háu stigi, tvær frá hverju landi: Íslandi, Danmörku, Noregi og Svíþjóð. Þessar átta kennslustundir voru greindar eigindlega út frá bæði samskiptum kennara við nemendur og kennslusniði.

Niðurstöðurnar sýna að athugunargildi fyrir hugræna virkjun voru lág í meirihluta kennslustundanna frá Íslandi. Tíminn í kennslustundunum var að mestu notaður til einstaklingsvinnu nemenda og útfærsla kennarans á verkefnum leiddi oft til þess að gildi voru lág. Greining stærðfræðilegu viðfangsefnanna sýndi að flest verkefni miðuðu að færni nemenda í að beita aðferðum á meðan tengsl við skilning á stærðfræðilegum hugtökum voru takmörkuð. Tengslin milli athugunargilda og skynjun nemenda sem mælinga á hugrænni virkjun voru veik. Dreifni í svörum nemenda var almennt meiri innan nemendahópa en á milli þeirra. Eigindleg greining á norrænu kennslustundunum sem metnar voru á háu stigi sýnir bæði fjölbreytt stærðfræðileg efnistök og kennslusnið. Hópvinna og bekkjarumræða voru fyrirferðarmikil snið, en allar kennslustundir innihéldu beina kennslu yfir stuttan tíma. Margar kennslustundir höfðu „hefðbundna" einstaklingsvinnu í stuttum sprettum. Kennslan í þessum kennslustundum einkenndist af áherslu á stærðfræðilegan skilning nemenda í gegnum tengsl stærðfræðilegra hugtaka, tíðar breytingar á tegundum samskipta, skýr hlutverk nemenda í skólastofunni til að stuðla að þátttöku þeirra, og endurgjöf til leiðsagnar.

Lykilorð:

Stærðfræðimenntun, gæði kennslu, hugræn virkjun, skólastofuathuganir, skynjun nemenda

Abstract

The doctoral research project underpinning this dissertation was aimed at developing a deeper understanding of cognitive activation in mathematics teaching in Iceland and in a Nordic context through classroom video observations and student perceptions. Ten mathematics teachers in Iceland participated. Three to four consecutive lessons in grade 8 were video-recorded. In total, 34 lessons in Iceland were analysed using the observation system PLATO. Mathematical tasks were identified and analysed using the Task Analysis Guide. The students, aged 13–14 (N = 217), responded to the Tripod student perception survey and their responses were compared to the PLATO observation scores. Specific lessons where cognitive activation scored high were selected for further analysis, two from each country: Iceland, Denmark, Norway, and Sweden. These eight lessons were analysed qualitatively in terms of teacher-student interactions and instructional format.

The findings show limited evidence of cognitive activation in a majority of mathematics lessons in Iceland. Lesson time was primarily used for students' individual work and the way teachers implemented tasks commonly resulted in low observation scores. The task analysis showed that most tasks were aimed at procedural fluency with limited connections to understanding mathematical concepts. The connection between the observation scores and student perceptions as indicators of cognitive activation was weak. Variance in student ratings was generally greater within classrooms than between them. The qualitative analysis of the outstanding Nordic mathematics lessons showed a variety of topics and instructional formats. Group-work and whole-class discussions were dominant, but all lessons included brief intervals of whole-class direct instruction. Many of these lessons had short sprints of "traditional" individual work. The teaching in these lessons was exemplified by an emphasis on student understanding through mathematical connection-making, frequent shifts between types of interactions, use of explicit student roles in the classroom to facilitate student engagement, and formative feedback.

Keywords:

Mathematics education, teaching quality, cognitive activation, classroom observation, student perceptions

Preface

This dissertation presents insights into teaching quality in mathematics lessons in Iceland and other Nordic countries. Many dimensions constitute teaching quality – the specific dimension in focus here is cognitive activation. Briefly put, cognitive activation is about how teachers offer students opportunities to engage with mathematically rich content in a way that can develop their understanding. This involves both the selection of tasks and the implementation of tasks during lessons. A lesson with high cognitive activation will typically include mathematically rich tasks or structured activity where students are invited to discuss, reason, and justify their solutions and results. Conversely, lessons with low cognitive activation will generally include rote tasks with a focus on producing correct answers and limited space for students' guided exploration.

A specific focus of this dissertation is the implementation of tasks. Teachers not only have to choose mathematically rich tasks – they also must implement these in a way that is conducive to develop student understanding of mathematical concepts and methods. Tasks with a high potential to develop mathematical understanding may be open to different solution paths, be conducive to a reasonable amount of productive struggle, and invite different explanations of students' thinking and reasoning. The way tasks are implemented can result in different classroom interactions. To explain the difference between interactions, the notion of students' accept/assert mode is useful (Mason & Johnston-Wilder, 2006). Giving students direct instructions for how to solve specific problems sets students in accept mode, to accept what the teacher says. However, where students are offered guidance to explore, make their own assertions and explain their solutions, they are in assert mode, with a richer opportunity to develop understanding.

Two main reasons inspired the formation of this research project. The first reason is based on my personal experience of the education system, from elementary student to mathematics educator. As a child, I was enthralled by mathematics. In first grade, all the kids in my class were to bring their favourite toy to school. Many boys brought action figures – I brought my abacus (and a picture from that day confirms this). However, as a teenager my interest and motivation for the subject waned substantially and grades went down. Fortunately, I found my passion for mathematics again as a young adult. During my years as a student in teacher education, I started questioning why my interest in mathematics dwindled – was it due to personal disposition, or was it associated with the quality of teaching that I received? Although this dissertation is not aimed at providing an answer to this personal question, it is important for me to mention. It has been and still is a driving force for my motivation in pursuing knowledge in the field of education and in the quality of mathematics teaching.

The second reason is based on my previous research on mathematical tasks. I was inspired by what I learned in my undergraduate studies about cognitively demanding tasks and chose to study them further in my master thesis. In an interview study of five upper secondary mathematics teachers, the aim was to develop knowledge of the views of Icelandic mathematics teachers toward cognitively demanding tasks for students in remedial courses. Despite most textbook tasks being low-demand, participating teachers generally expressed a more positive view toward using high-demand tasks in their lessons (Sigurjónsson & Kristinsdóttir, 2018). The study was limited to teachers' views toward different tasks and did not include systematic observation and analysis of how teachers implemented tasks in lessons. Teachers make decisions in how they implement tasks in their teaching that are important for student learning. I found it critical to learn more about how mathematical tasks unfold in the classroom.

In this context, the University of Iceland's participation in the Quality in Nordic Teaching (QUINT) project from 2018 came to be at a fortunate time for me. The QUINT ambition is to produce new insights into what characterises teaching quality in Nordic classrooms. When I learned of a QUINT PhD candidate position for research on mathematics teaching, I immediately saw the opportunity to continue the work I had started – to inquire further about the realities of mathematics teaching with video-based classroom observations and student perceptions. My contribution to the project was based on my participation in the data collection and joint analyses in Iceland, adding to the unique QUINT video library – and ultimately, this article-based doctoral dissertation. My PhD fellowship with QUINT, providing access to infrastructure and affiliation with experienced scholars, was pivotal in continuing my pursuit of knowing more about quality in mathematics teaching. This dissertation is the result of that work.

Contents

List of Abbreviations

Abbreviation	Meaning
CA	Cognitive Activation
CD	Classroom Discourse (PLATO element)
COACTIV	Professional Competence of Teachers, Cognitively Activating Instruction, and the Development of Students' Mathematical Literacy (German research program)
GTI	OECD's Global Teaching InSights (aka TALIS Video Study)
IC	Intellectual Challenge (PLATO element)
IRE	Initiation-Response-Evaluation interaction pattern
LPS	The Learner's Perspective Study
LISA	Linking Instruction and Student Achievement study
MET	Measures of Effective Teaching study
OECD	Organisation for Economic Co-operation and Development
PISA	OECD's Programme for International Student Assessment
PLATO	Protocol for Language Arts Teaching Observations
QUINT	Quality in Nordic Teaching research centre
TAG	The Task Analysis Guide
TALIS	OECD's Teaching and Learning International Survey (GTI)
TIMSS	The Third International Mathematics and Science Study

List of Figures

List of Tables

List of Original Papers

This thesis is based on the following original publications, which are referred to in the text by their Roman numerals:

I. Sigurjónsson, J. Ö., & Gísladóttir, B. (2020). Vitsmunaleg áskorun í stærðfræðikennslu á unglingastigi [Intellectual challenge in mathematics teaching in lower secondary schools]. *Icelandic Journal of Education*, *29*(2), 149–172.

II. Sigurjónsson, J. Ö., Sigurðardóttir, A. K., Gísladóttir, B., & van Bommel, J. (2022). Connecting student perceptions and classroom observations as measures of cognitive activation. *Nordic Studies in Education, 42*(4), 1–19.

III. Sigurjónsson, J. Ö. (accepted for publication). Teaching mathematics with high cognitive activation: Characteristics of outstanding lessons. *Nordic Studies in Mathematics Education*.

All papers are reprinted by kind permission of the publishers.

List of Related Publications

As a doctoral student I have authored or co-authored the following publications:

- Pálsdóttir, G., & Sigurjónsson, J. Ö. (2021). Norræn ráðstefna um rannsóknir á sviði stærðfræðimenntunar [A Nordic conference on mathematics education research]. *Flatarmál*, *28*(1), 33–35.

- Fauskanger, J., Naalsund, M., Nilsen, H. K., Nortvedt, G. A., Pálsdóttir, G., Portaankorva-Koivisto, P., Radišić, J., Sigurjónsson, J. Ö., Wernberg, A. (2022). Bringing Nordic mathematics education into the future - An analysis of the themes and impact of papers presented at the NORMA20 conference. In Nortvedt, G., ... Wernberg, A. (Eds.) *Proceedings of Norma 20, the ninth Nordic Conference on Mathematics Education* (pp. 281–294).

- Sigurjónsson, J. Ö. (2022). Cognitively activating mathematics lessons: a Nordic comparative study. In L. Mattsson, J. Häggström, M. Carlsen, C. Kilhamn, H. Palmér, M. Perez, & K. Pettersson (Eds.), *The relation between mathematics education research and teachers' professional development. Proceedings of MADIF13. The thirteenth research seminar of the Swedish Society for Research in Mathematics Education, Växjö, 2022* (p. 138). SMDF.

- Sigurjónsson, J. Ö., Kristinsdóttir, B., & Gíslason, I. (in press). Evrópsk ráðstefna um rannsóknir á sviði stærðfræðimenntunar [A European conference on mathematics education research]. *Flatarmál*, *29*(1).

- Sigurjónsson, J. Ö. (in press). Cognitively activating lessons: A Nordic comparative study. *CERME12 proceedings*.

Declaration of Contribution

Paper I: Intellectual challenge in mathematics teaching in Icelandic lower seondary schools

I participated in collection of data as part of QUINT's data collection in Iceland. I identified all the tasks used in the lessons from videos and data logs. I analysed all the task data and participated in coding the video data (around one third) as well as double-coding with other PLATO raters. I wrote the paper along with my supervisor Berglind. She and Sólveig Zophoníasdóttir coded the rest of the video data. I discussed the findings extensively with both of my supervisors, Berglind and Anna Kristín. The article was published in Icelandic but was translated to English by me for this dissertation.

Paper II: Connecting student perceptions and classroom observations as measures of cognitive activation

I conducted the analysis and wrote the paper. Data collection and video data analysis was the same as in Paper I. The findings were discussed extensively with my whole committee: Anna Kristín, Berglind, and Jorryt. The discussions and feedback significantly shaped analysis of the data, the structure of the paper and presentation of results.

Paper III: Teaching mathematics with high cognitive activation: Characteristics of outstanding lessons

I designed the selection criteria, analysed the data, and wrote the paper. Data collection, translation and transcription was done through colleagues in the QUINT centre in Denmark, Iceland, Norway, and Sweden. I had productive discussions and feedback on my tentative findings at seminars and meetings with Jorryt and her doctoral student Jimmy Karlsson during my stay in Karlstad, Sweden, and at a seminar with the QUINT Iceland group.

Acknowledgements

I appreciate you, the reader, for showing interest in my work. If you would like to reach out with any questions or comments, my email address is johann.orn@outlook.com.

First, I would like to express gratitude to all the Nordic teachers and students who participated in the study by agreeing to have their lessons video-recorded and analysed. It is not self-evident to offer such professional vulnerability in the pursuit of scientific and educational development. Thank you for your invaluable contribution to the research.

I am both grateful and fortunate in having Anna Kristín Sigurðardóttir and Berglind Gísladóttir as my supervisors. From the beginning they offered me both solid guidance as well as trust in working independently. They helped me navigate the jungles of academia, suggesting avenues to grow through participation in seminars, conferences, and courses, providing constructive critique on my work, seeking funding, and generally providing me with opportunities for educative experiences during my doctoral studies. Sincere thanks to them both: Kærar þakkir.

Jorryt van Bommel, expert on my doctoral committee: Special thanks for always being ready to provide superb feedback and new perspectives. Your professional insight was very valuable to me throughout the project. I owe my thanks to you for making my stay in Karlstad so successful. Honourable mentions to Jimmy Karlsson and Marcus Gustafsson; Tack så mycket!

This dissertation reports on my research project which is based on my fellowship with Quality in Nordic Teaching (QUINT). The QUINT research centre has been a vital resource for me by offering partnership in collection and use of data, regular seminars, courses relevant to the doctoral project, participation in conferences and doctoral seminars, and cooperation with other scholars and doctoral students alike, both within and outside the centre. To my fellow doctoral students within the QUINT centre: Alexander, Anna, Cæcilie, Jenny, Jonas, Peter, and Valgarður. Thank you for the mutual support, despite all the challenges! Honourable mentions also to Jennifer Maria Luoto and Roar Bakken Stovner for cooperation and conversation on approaches to the mathematics classroom data. I must specifically mention the centre director, Kirsti Klette, for the opportunities and experiences she provided for me and us all, as well as post-doc Mark White for his insight and assistance. To all my QUINT colleagues: Takk skal du ha, tack så mycket, tak skal du have, kiitos, takk fyrir, thank you!

For translation and transcript work of non-Icelandic data, I would like to personally acknowledge Alexander Selling (University of Oslo), for translating the selected lessons from Norway and Sweden, and Natasha Sterup (University of Southern Denmark) for translating lessons from Denmark. Not only did they translate and transcribe the entire lessons but they also helped greatly in validating my interpretation of contextual issues. Quite literally, I could not have done the analyses for Paper III without their contribution.

I would like to acknowledge my colleagues in the School of Education at the University of Iceland. I especially thank my fellow doctoral candidates for their companionship and support – too many to mention in full, but they know who they are! Extra special thanks to: Guðbjörg Pálsdóttir for multiple borrowed books and collaboration for the NORMA conference; Jónína Vala Kristinsdóttir for further multiple borrowed books and providing information and opportunities to me as master thesis supervisor; Annadís Greta Rúdólfsdóttir for her helpful advice (and even more books) before the defence; Friðrik Diego for his companionship, Friday riddles and inviting me to contribute to the maths competition committee; and Freyja Hreinsdóttir and Ólöf Steinþórsdóttir for their valuable and thorough feedback on my doctoral project's interim evaluation.

For cooperation in publishing, and permission to include papers I, II, and III in the dissertation, I thank the following: Icelandic Journal of Education, Nordic Studies in Education, and Nordic Studies in Mathematics Education.

I am very grateful for the financial support provided by the following funds and associations which were fundamental for being able to devote enough time to my doctoral project in the past years: The Icelandic Research Fund (Rannsóknasjóður Grant 206754-051 through University of Iceland, School of Education), QUINT (Nordforsk Grant 87663 through University of Oslo, Department of Teacher Education and School Research, and the QUINT mobility fund), University of Iceland's Doctoral Student Travel Fund (Ferðasjóður doktorsnema), and the Steingrímur Arason Memorial Fund (Styrktarsjóður Steingríms Arasonar).

Last, but certainly not least, to my friends and family: My heartfelt thanks for you all and your support throughout this journey. To my parents, Sigrún and Sigurjón, for being by my side in everlasting support and belief in me. To my fiancée, Sólrún, for her endless support, good humour, shared adventures, and constant readiness to turn my questionable visual design into good design. To Skotta, for being such a good cat. To each and every one else: Thank you for being there, always.

1 Introduction

If one would ask a few persons at random from anywhere in the world *what characterises good teaching*, the responses would vary based on a range of factors. Perhaps they base their responses on experiences and memories they have from their years in school, different personalities of certain teachers that they looked up to, or otherwise relate to the cultural context in which their conception of teaching originates. Perhaps they themselves have received teacher training, have instructed their co-workers on work-related tasks, or have tutored their friends or family members. Regardless of differences in responses, most would agree that the quality of teaching matters – and the academic literature is in support of that claim (Cochran-Smith, 2003).

But what is the exact meaning of *teaching quality*[a]? Educational researchers have worked on disentangling this concept in various ways, resulting in different theoretical conceptualisations and operationalisations. There is general agreement that teaching quality is a multi-dimensional construct, i.e., it is not measured as a single metric, but consists of different domains or dimensions (Croninger et al., 2012). Some dimensions are more generic, and others are more specific to the subject matter. An example of a generic framework is the German framework of Three Basic Dimensions: Student Support, Classroom Management and Cognitive Activation (Praetorius et al., 2018). Student support refers to teaching being sensitive to individual needs to support student motivation. Classroom Management refers to teaching that minimises interruptions and monitors student attention. Cognitive Activation, a central concept in this research project, refers to the selection of adequately challenging tasks which may be used to engage students in higher order thinking to develop understanding. Although many conceptualisations and approaches to the meaning of teaching quality and its various dimensions exist, the focus of this research project is on the dimension of cognitive activation in mathematics lessons.

1.1 Background

The origins of the term "teaching quality" in educational research can be traced back to an article by Carroll (1963). From an era where behaviourism was the dominant paradigm in educational research, Carroll suggested a conceptual model of school learning as a function of time with two external factors: opportunity (as "time allowed for learning") and teaching quality. Teaching quality was remarked as "the most elusive

---

[a] Teaching quality is sometimes called instructional quality, or phrased in the literature as teaching effectiveness – I have chosen to use the term *teaching quality*.

quantity" of the model, and that measurements of it were at the time "practically non-existent". Yet, he regarded teaching quality as "one of the most important manipulable variables in educational psychology" (p. 727).

To understand and measure teaching quality, educational researchers have in recent years and decades developed various instruments, including student perception surveys and observation systems (Bell et al., 2019; Wallace et al., 2016). Improvements in video technology have allowed researchers to rewatch and re-evaluate classroom video data and further refine observation systems. In an analysis of twelve observation systems in mathematics, Praetorius and Charalambous (2018) found cognitive activation among the most prominent dimensions measured. They concluded that the field would benefit from more collaborative work and agreed-upon standards for studying teaching quality. Building on a metaphor by Gitomer (2009) of trying to detect signals in a noisy environment as an alarm clock in a subway, synchronising attempts to operationalise and measure teaching quality "would be as if we concurrently use several alarms in the subway. If all of them are appropriately synchronised [...] the signal is going to come out louder" (Praetorius & Charalambous, 2018, p. 551).

Cognitive activation involves both the selection and implementation of challenging tasks. Mathematical tasks – solving problems – is a fundamental part of mathematics teaching practices. In Pólya's words, "solving problems can be regarded as the most characteristically human activity" (1966, p. 126). Pólya differentiated between routine and non-routine problems in the sense that non-routine problems demand some degree of creativity and originality from the student, while the routine problems do not (Pólya, 1981). However, to select and assign non-routine tasks does not suffice – the implementation of tasks during lessons is key in offering students opportunities to develop understanding, i.e., to be cognitively activated. Evidence suggests a positive connection between cognitive activation and both student achievement (Klieme et al., 2001; Krauss et al., 2020) and enjoyment (Cantley et al., 2017; Ekatushabe et al., 2022).

Mathematics teaching in Iceland has been described in similar ways as in the other Nordic countries. Studies of mathematics teaching in Iceland have found lesson time to be mostly used for individual work in textbooks with few examples of group-work and discussions (Gunnarsdóttir & Pálsdóttir, 2015; Þórðardóttir & Hermannsson, 2012). In Sweden, lesson time in mathematics also seems dominated by individual work and commonly low cognitive activation, indicating a focus on procedural competency (Boesen et al., 2014; Tengberg et al., 2021). Studies of mathematics teaching in Norway have painted a similar picture, with as low as 5% of lesson time consisting of group-work and the main focus of teacher feedback being on students' procedural skills (Bergem & Pepin, 2013; Stovner & Klette, 2022). However, no studies of Nordic mathematics classrooms have explicitly studied cognitive activation as a dimension of teaching quality.

1.2 Originality and rationale

The main originality of the study is in its novel theoretical and methodological approach to empirically explore Nordic mathematics teaching, i.e., to systematically study cognitive activation as a dimension of teaching quality in Iceland and in a Nordic context, using both video-recordings of lessons and student surveys. Video data have rarely been used in studying mathematics teaching in Iceland and never on the scale as in this study. Nevertheless, video-based classroom research has some history. Large-scale video-based research in mathematics classrooms has been conducted since the 1990s, with some notable international comparative research being The Third International Mathematics and Science Study (TIMSS), The Learner's Perspective Study (LPS) and VIDEOMAT (Clarke, Emanuelsson, et al., 2006; Jacobs et al., 2003; Kilhamn et al., 2018). Both TIMSS and LPS included countries from America, Asia, and Europe, while VIDEOMAT had a Nordic perspective in comparison to the United States (California). Iceland has not participated in a large-scale classroom video study before.

Cognitive activation has been researched in different ways, such as task analysis and classroom observation (e.g., Neubrand et al., 2013; von Kotzebue et al., 2020). The cognitive demand of tasks has long been a topic of research in mathematics education (Stein & Smith, 1998; Tekkumru-Kisa et al., 2020). Although connections between cognitively demanding tasks and student learning are widely acknowledged, a citation analysis concluded that claims about a direct connection have sometimes been overstated (Otten et al., 2017; Stein et al., 2009). The implementation of rich tasks through cognitively activating teaching practices has been suggested as a mediating factor, where tasks alone offer *objective* cognitive activation potential, while through interactions in the classroom, teachers offer the *implemented* cognitive activation potential (Helmke, 2015; Weingartner, 2021). Research on cognitive activation in mathematics lessons has been largely conducted within a German or North-American context (see e.g., Baumert et al., 2013; Tekkumru-Kisa et al., 2020). While some studies have found cognitive activation to be positively linked with student achievement gains (Klieme et al., 2001; Krauss et al., 2020; Lipowsky et al., 2009), research is scarce in developing empirical understandings of how teachers engage in teacher-student interactions that create potential for cognitive activation.

Cognitive activation has not been specifically researched in mathematics teaching in Iceland before. While studies have found some promising results between student perceptions of teaching and student outcomes (e.g., Sandilos et al., 2019), the connection between student ratings and classroom observation scores remains somewhat unclear (Schweig, 2014; Wallace et al., 2016). By exploring these interactions in a Nordic context, a novel empirical contribution is made in two ways: firstly, by the context in which the study is conducted, and secondly in creating the necessary empirical knowledge useful to apply research results to teacher training and professional development in a Nordic context.

Moreover, to aid in applying research results to further develop teacher education and teaching practice, there is a need to develop empirical understandings of lessons with high cognitive activation. I argue that the results of this dissertation can have implications for policy and practice alike, as it can aid in making informed choices about the development of teacher education and teaching practice. It is also of relevance for the research community, as studying cognitive activation in mathematics teaching in a Nordic context can shed light on the affordances and constraints of applying research frameworks in different cultural contexts. Further, there is a need to develop empirical understandings of how teachers enact cognitively activating practices in interaction with students during lessons. The implementation of tasks in lessons is fundamental to the learning opportunities that are created.

1.3 Aims

The overarching aim of the research project was to develop a deeper understanding of the teaching quality dimension of cognitive activation in mathematics in Iceland and in a Nordic context. The overarching aim can be broken down into sub-aims:

1. To assess the cognitive activation potential of mathematical tasks in lower secondary classrooms in Iceland.

2. To assess the level of cognitive activation in mathematics teaching in lower secondary classrooms in Iceland.

3. To examine student perceptions of cognitive activation.

4. To examine the connection between observed level of cognitive activation and student perceptions.

5. To enrich empirical understandings of instructional formats in Nordic mathematics lessons considered cognitively activating.

6. To enrich empirical understandings of teacher-student interactions in Nordic mathematics lessons considered cognitively activating.

1.4 Research process

The research project started in 2019 and took around three-and-a-half years in total to complete. In the first year, the very first task was to receive instructions and training in data collection for the QUINT research initiative. This allowed participation in collecting data by video-recording lessons and administering a survey in a couple of schools in Iceland. The next step was to receive training in the observation system to analyse the mathematics classroom video data collected in Iceland. The second year consisted of employing complementing methods of analyses to these data. Firstly, mathematical tasks were identified from the video-recorded lessons and analysed. Secondly, student perceptions from the survey were analysed in connection to the classroom observation data. In the third year, the primary task was to consider QUINT video data available from the other Nordic countries. A further qualitative analysis of specific Nordic mathematics lessons with outstanding observation scores was the final analytical task. Writing of academic papers and contributions at conferences took place between the different analyses and toward the end of the research process.

Three academic papers were written as part of the research project. The first paper was written in Icelandic with one co-author (and co-supervisor) also involved in the QUINT project. It was published in the *Icelandic Journal of Education*. The English translation of this article is included in this dissertation. The second paper was written in English with three co-authors, all involved in the aforementioned project as well as being supervisors and members on my doctoral committee. It was published in *Nordic Studies in Education*. The third paper was written in English with no co-authors. It has been accepted for publication in *Nordic Studies in Mathematics Education*[b]. I am the first author on all three papers.

1.5 Dissertation outline

This dissertation contains five chapters. This first chapter has been an introduction to the background of the doctoral research project, its originality and aims. Chapter 2 is a review of literature that aims to contextualise the research project to its Icelandic and Nordic context, present theoretical perspectives to outline the conceptual frameworks of teaching quality that the project builds on and its methodological background. Chapter 3 aims to explain the research design and the research methods used in the study to reach the research aims. Chapter 4 contains a presentation and extended overview of findings in relation to the aims and research questions. Finally, chapter 5 concludes the dissertation by a discussion of its contribution to empirical, theoretical, and methodological development, with reflections, suggestions for future research, and concluding remarks.

[b] The submitted manuscript is included in this dissertation. The published version of the article will be subject to changes from the submitted manuscript.

2 Literature review

This chapter provides a review of literature in three sub-chapters. Firstly, the Icelandic and Nordic research context is explained with reference to previous studies on Nordic mathematics teaching, as well as a historical context of video-based research in mathematics education. Secondly, theoretical perspectives are presented where cognitive activation is the central focus, as well as perspectives of interaction research in the mathematics classroom. Thirdly, previous literature is discussed related to methodology, i.e., classroom observation systems and student perceptions of teaching quality.

2.1 Research context

Teaching and learning are highly context-sensitive, which means they are to a large extent influenced by factors such as social norms and ideals. As this is a research project on Nordic teaching, it is important to communicate in detail the Nordic context in which the research is situated. In this sub-chapter, the local context of Iceland and the Nordic countries is explained with reference to relevant literature on mathematics teaching, as well as the historical context of video-based research in mathematics education.

2.1.1 Mathematics teaching in Iceland

Iceland is an island nation in the North Atlantic with about 370,000 inhabitants. The Icelandic school system requires children aged 6-16 to attend compulsory school ("Lög um grunnskóla nr. 91/2008 [Law on compulsory schools, no. 91 of 2008]", 2008). Grades 8–10 are referred to as lower secondary school ("unglingastig"), with students aged 14–16. Each grade level cohort in Iceland consists of around 4,500 to 5,000 students (Statistics Iceland, 2020). Since 2010, acquiring a teaching certificate requires having completed the equivalent of five years of university education, including a master's degree in education ("Lög um menntun, hæfni og ráðningu kennara og skólastjórnenda við leikskóla grunnskóla og framhaldsskóla nr. 95/2019 [Law on education, competence an employment of teachers and school administrators at pre-schools, compulsory schools and upper secondary schools, number 95 of 2019]", 2019).

Mathematics teaching in Iceland has been criticised in recent years, where homogenous teaching methods, a narrow view of the curriculum and a negative trend in PISA results have been some points of critique (OECD, 2016, 2019; Óskarsdóttir, 2014). In a report involving eight lower secondary schools, students' individual work in textbooks with assistance from the teacher was the most common practice (Þórðardóttir & Hermannsson, 2012). A total of 83% of teachers reported this instructional format to be incorporated either often or very often. Among lessons observed for the report, about a third included direct whole class instruction and student discussions. A report on upper secondary mathematics teaching described a similar trend, adding that tasks where students have to reason mathematically to develop conceptual understanding seemed virtually non-existent (Jónsdóttir et al., 2014). In an analysis of directly observed 51 lower secondary mathematics lessons, Gunnarsdóttir and Pálsdóttir (2015) came to the same conclusion: in most mathematics lessons in Iceland, the students work individually in textbooks with the teacher walking between desks interacting with students – although there were some examples of instructional practices that emphasised group-work and discussions.

What studies of mathematics teaching in Iceland have found is perhaps unsurprising in light of studies on prospective teachers in Iceland. A study building on the *Mathematical knowledge for teaching* (MKT) measures indicated that prospective teachers in Iceland exhibit largely procedural and algorithmic knowledge of mathematics and that they experience difficulty in evaluating alternative solution methods and working with fractions (Jóhannsdóttir & Gísladóttir, 2014). A pre-assessment of first-year students in teacher education programs averaged 44% correct answers to multiple-choice questions on compulsory mathematics content (Hreinsdóttir & Diego, 2019). Furthermore, a recent study exposed vastly negative life experiences and dispositions toward mathematics among prospective teachers (Gíslason & Gísladóttir, 2021). Negative dispositions and poor content knowledge among prospective teachers do not suggest good preconditions for high quality teaching in mathematics.

Iceland has participated in very few comparative studies of mathematics teaching. In a small-scale video study, the only one of its kind comparing Icelandic mathematics teachers to teachers from other countries, Savola identified Finnish mathematics teachers as "rather traditional", while describing Icelandic teachers as "progressive-minded" and mainly using learner-based strategies (Savola, 2010). A major contrast was found in the independent student learning in Iceland compared to the emphasis on whole-class interaction in Finland. He concluded that many Icelandic teachers emphasised individualisation and learner control at the cost of content-related discourse and reasoning. These results raise questions about to what extent these differences account for different results in international measurements such as PISA (OECD, 2019). Suggestions for improvement have included observing the case of Sweden where PISA results trended upward following a nation-wide professional development program (Hreinsdóttir, 2019). When exploring avenues for development and improvement, it is rational to look to practices in neighbouring countries that indicate more favourable outcomes.

2.1.2 Mathematics teaching in the Nordic countries

The Nordic countries share a history of collaboration in the field of educational research since at least the 1970s with the foundation of the Nordic Educational Research Association (NERA). Nevertheless, mathematics education is considered a young research field in a Nordic context (Fauskanger et al., 2022). As such, the research literature on mathematics teaching in the Nordic countries has grown rapidly in the past 15-20 years, mostly by empirical studies on a range of different research topics (Grevholm, 2021; Rønning, 2019).

The school systems in the Nordic countries share many similarities. Excluding Denmark, a master's degree is generally required for a teaching certificate in the Nordic countries. The Nordic model of education is based on the "school for all" ideal, i.e., that schools should be inclusive, comprehensive, non-tracked and providing easy passages between school levels (Blossing et al., 2014; Klette et al., 2021). The compulsory level structure in Iceland is in principle in accordance with the Nordic model ideals of a school for all (Sigurðardóttir et al., 2014). Thus, Nordic classrooms share considerable structural similarities but also distinct differences, and for this reason they are claimed to represent an ideal context for comparative ambitions (Dahl & Stedøy, 2004; Klette et al., 2017).

Research on mathematics teaching in Sweden has indicated a procedural focus and low cognitive activation. A study on competence reform in Sweden revealed that even fifteen years after its initiation, classroom practices were still mostly focused on developing competency in carrying out procedures (Boesen et al., 2014). Some scholars went as far as describing mathematics teaching as students usually working with their textbooks at their own pace "without any teaching" (Pehkonen, Hemmi, et al., 2018). More recently, a nation-wide mathematics professional development programme *Boost for Mathematics* (*Matematiklyftet*) was launched with the aim to develop teaching culture through collegial teacher learning (Österholm et al., 2021). An increase in Sweden's PISA score in mathematics followed the programme, though direct causality is not claimed (Hreinsdóttir, 2019; OECD, 2019). A recent video study on teaching quality in Sweden showed around half of observed segments focused on student's individual work, with around 30% being mostly whole-class instruction, and the rest (~20%) either pair work or group-work. Around half of the segments were scored on the low end for intellectual challenge, and 77% were on the low end for classroom discourse, indicating low cognitive activation (Tengberg et al., 2021).

Research on mathematics teaching in Norway has painted a similar picture. An analysis of 172 lower secondary mathematics lessons found that teachers used considerable lesson time giving feedback that was more commonly focused on procedural skills rather than conceptual understanding (Stovner & Klette, 2022). A study of 38 mathematics lessons in grade 9 found 95% of lesson time was used for individual seatwork and whole-class instruction, quite evenly distributed, meaning only 5% of lesson time consisted of student group-work (Bergem & Pepin, 2013). Although a subset of teachers does offer students to participate in mathematical discourse focused on development of conceptual understanding, the broad picture shows a strong focus on procedures.

Descriptive studies in Denmark and Finland take a somewhat different view. A small-scale study described Danish mathematics lessons as often balancing exploratory and more instructional work, identifying two "lesson types": one with the teacher assuming the role of instructor-in-dialogue on procedural mathematics with guided and interactive exploration, and the other with problem-oriented group explorations using an open questioning approach interspersed with episodes of whole-class instruction-in-dialogue (Kelly et al., 2013). In a Finnish study, mathematics teaching was described as both "teacher-centred" and reliant on textbooks to help maintain a high quality in teaching (Pehkonen, Piht, et al., 2018). There is reason to doubt that the reliance on mathematics textbooks in Finland is of the same kind as reported in Iceland. In Savola's small-scale video study (2010), he identified Finnish mathematics teachers as "rather traditional" in emphasising whole-class interaction. The contrast to the emphasis on individualisation and learner control in Iceland was argued to be at the cost of content-related discourse and reasoning. Although these descriptive studies were small-scale and by no means encapsulate any "national pattern", they give some insight into what teaching looks like in these contexts. Looking for further explanations of differences in teaching practices and student outcomes, it is useful to look to previous large-scale video studies of mathematics teaching.

2.1.3 Video-based research in mathematics education

The use of video as a method of generating data for research has considerable history. Utilisation of video technology in the field of social science research can be dated as far back as 1898, with anthropologist Alfred C. Haddon recognising moving images "as a resource for the analysis and presentation of cultural practices" (Heath, Hindmarsh, & Luff, 2010). Over 100 years later however, Heath, Hindmarsh & Luff go on describing the use of video as an investigative tool of human activity as "neglected" within the social sciences – despite the emergence of use of visual media, particularly within social anthropology and so-called "workplace studies". Their concerns echo those voiced in the 1970s by Margaret Mead, who claimed that "the behaviour that film could have caught and preserved for centuries [...] disappears right in front of everybody's eyes." (Mead, 1974). Video-based research in mathematics education started to gain traction in the 1990s, with large-scale studies such as the Third International Mathematics and Science Study (TIMSS) and The Learner's Perspective Study (LPS) being conducted at an international level (Clarke, Emanuelsson, et al., 2006; Stigler et al., 1999). Their history and main findings give context to the current study.

The first TIMSS video study was conducted in 1995, where over 200 eighth-grade mathematics lessons in Germany, Japan and the United Sates were randomly sampled. The lessons were filmed with a single camera pointed at the teacher and were supplemented by a post-lesson teacher questionnaire. In spirit of its pioneering nature within the field of mathematics education, its primary goals were not only to provide rich descriptions of teaching in these different countries but also to "assess the feasibility of applying videotape methodology in future wider-scale national and international surveys of classroom instructional practices" (Stigler et al., 1999). They found both advantages and disadvantages of using video in classroom research. One of the main advantages, in their view, was how video enabled the study of complex processes, while disadvantages included issues of teacher sampling. In their attempt to determine so-called "national patterns of teaching", one of the main results was the vastly different approach to teacher development found in Japan compared to Germany and the USA, for example with the use of "lesson study groups" (Stigler & Hiebert, 1997).

The second TIMSS video study, conducted in 1999 and sometimes labelled TIMSS-R, enlarged the scope by incorporating a total of seven participating countries: Australia, the Czech Republic, Hong Kong SAR, Japan, the Netherlands, Switzerland, and the United States. This follow-up study added a second camera pointed in the opposite direction, toward "the back" of the classroom. Other data types were also collected, including both teacher and student questionnaires, samples of student work, textbook pages, and sample tests. In addition to the goals of the TIMSS 1995 video study, TIMSS 1999 aimed to develop "methods for communicating the results of the study, through written reports and video cases, for both research and professional development purposes" (Hiebert et al., 2003). The study found that while many similarities were found across countries, the differences raised more questions than answers, highlighting the complexity of teaching (Hiebert et al., 2003).

In a reflection on both TIMSS video studies, Jacobs, Hollingsworth & Givvin (2007) addressed a serious pitfall in not collecting enough supporting data for contextualisation purposes. They suggest that it is wise to collect as much supplementary data as resources allow, even if the original vision of the research project may only make partial use of it. Furthermore, in a report on the TIMSS 1999 video study, capturing single lessons as opposed to a sequence of lessons was described as a major limitation (Lokan et al., 1999). Following the TIMSS video studies, the Board on International Studies in Education specifically recommended pursuing research projects of different scopes and sizes that appropriately use video technology as a research tool. It further recommended exploring the creation of video archives for international comparative research in education, similar to the work that had started with the TIMSS studies (Lokan et al., 1999; National Research Council, 2001).

Before those recommendations were published, a larger international study was already underway. The Learner's Perspective Study (LPS) compared mathematics classrooms in twelve countries from around the globe: Australia, China (Hong Kong, Shanghai and

Macau), the Czech Republic, Germany, Israel, Japan, Korea, the Philippines, Singapore, South Africa, Sweden, and the USA (Clarke, Emanuelsson, et al., 2006). The LPS used three video cameras to simultaneously document sequences of at least ten lessons with teachers deemed highly competent by local criteria, with three participant teachers from each country. Drawing further from the pitfalls of the TIMSS studies, they supplemented the video data with student written material, researcher field notes, test, and questionnaire data, in addition to both teacher and student interviews. The analysis of these massive and diverse datasets was undertaken from a variety of theoretical perspectives, seen as complementary as opposed to conflicting, since both aspired to inform practice and advance theories in different ways (Clarke et al., 2006). For instance, in Sweden and Hong Kong, variation theory (Marton & Booth, 1997) was employed to compare classroom practices in Shanghai, while in Australia the same lessons were analysed in relation to the distribution of the responsibility of knowledge generation (Clarke & Xu, 2008). The main results highlighted just how culturally-situated classroom practices around the world are, with patterns of participation "reflecting individual, societal and cultural priorities and associated value systems" (Clarke et al., 2006).

Both TIMSS and the LPS have been criticised for overgeneralising about differences between countries (Xu & Clarke, 2019). One example is the case of Japan. None of the lessons from Japan in the LPS data matched the "national Japanese model" that was described based on the TIMSS data (Stigler & Hiebert, 1999). Further, the Japan lessons were not considered to fit any "Asian" stereotype, and the popular pedagogy in Hong Kong SAR more resembled the "German model" reported in TIMSS, with careful explanations and a dominant "transmission mode" (Lopez-Real et al., 2004). Inferring any continental or national "teaching model" by generalising research results to countries or larger regions remains fraught with difficulties.

In the early 2010s the VIDEOMAT study was conducted, where video data from algebra lessons in Finland, Sweden, Norway, and the USA were analysed to identify and compare instructional strategies, classroom interaction, and pupil reasoning. In a second phase, participant teachers discussed their practices in focus group sessions where the aim was to develop their practice by viewing their own instruction (Kilhamn et al., 2018). In VIDEOMAT, three cameras were used, similar to the LPS approach. Both large-scale and small-scale studies in education are increasingly using video as a data source, with the Global Teaching InSights (GTI, initially referred to as the TALIS Video Study) being a recent and significant large-scale example (OECD, 2018, 2020). The ambition of the QUINT research centre has been to move the lens toward the Nordic countries in a continuation of this global development. A comparison of different aspects of the large-scale video-based studies discussed is presented in **Table 1**, comparing the number of participating countries, number of cameras used, number of lessons recorded, and the number lessons of sequential lessons collected from each participating teacher. The mathematics part of QUINT's video-based LISA Nordic study is added in the right-most column.

Table 1. Aspects of large-scale video-based studies in mathematics education.

	TIMSS 1995	TIMSS 1999	LPS 2000	VIDEOMAT 2011-2014	GTI 2017-2018	QUINT 2017-2020
Number of countries	3	7	12	4	8	5
Number of cameras	1	2	3	3	1	2
Number of lessons	231	638	≈360	≈90	≈1360	≈160[c]
Sequence of lessons	1	1	10	5	2	4
Content	Various	Various	Various	Algebra	Quadratic equations	Various

2.2 Theoretical perspectives

Teaching quality can be viewed from multiple theoretical perspectives. The aim of this chapter is to delineate the theoretical perspective taken in this dissertation and to justify the conceptual frameworks employed. Empirical questions about teaching quality should be addressed with careful definitions and justifications. This chapter begins by explaining cognitive activation as the teaching quality dimension in focus in this dissertation. It is explained both in terms of the cognitive activation potential of tasks, and cognitive activation as facilitation by classroom interactions. Lastly, research and theories on interactions in the mathematics classroom are discussed.

2.2.1 Cognitive activation as a dimension of teaching quality

In the constructivist sense, a central aim of teaching is to develop students' conceptual understanding (Glasersfeld, 1995). To understand mathematical concepts and their relation to other concepts and procedures requires students to rebuild existing knowledge by actively engaging with content, thinking reflectively, and engaging in a reasonable amount of productive struggle (Hiebert & Grouws, 2007; Kunter & Voss,

[c] If lesson data collected in language arts and social science were also counted, the total number in the QUINT data collection is 560 lessons (Klette, 2022).

2013). Student understanding is widely formulated as an educational goal (Newton, 2000). Diederich and Tenorth proposed three generic educational goals – student attentiveness, student motivation, and student understanding – which built on various European and other Western traditions in educational science, including the work of John Dewey (Diederich & Tenorth, 1997; Praetorius et al., 2018). Cognitive activation is a concept originally coined by Klieme, Schümer and Knoll (2001) that describes to what degree teachers address the educational goal of student understanding. Teachers can offer a potential for students to actively develop their understanding in several different ways, e.g., select and pose challenging tasks, invite students to participate in classroom discourse ("Socratic teaching"), or support metacognition by reflective exercises, to name a few (Klieme et al., 2006).

A growing amount of evidence suggests a positive connection between cognitive activation potential and student outcomes. Cognitive activation has been found to significantly impact student achievement in a German context (Kunter et al., 2013; Lipowsky et al., 2009). Studies have found cognitive activation to be positively related to student enjoyment and interest in mathematics, with some evidence suggesting the possibility of being an emancipatory force in especially boosting girls' enjoyment (Cantley et al., 2017; Lazarides & Buchholz, 2019). However, time constraints may hinder teachers from implementing cognitive activation strategies (Teig et al., 2019). Furthermore, cognitive activation is significantly predicted by teachers' pedagogical content knowledge (Baumert et al., 2010; Krauss et al., 2020). This suggests that prospective teachers may need specific support in making time to incorporate and develop cognitive activation strategies as part of their teaching repertoire.

Cognitive activation can be and has been researched from different perspectives. In an analysis of 12 classroom observation frameworks, cognitive activation was identified in each framework according to a definition in terms of three teaching practices: (1) a) selection of challenging tasks, and b) use of mathematically rich practices, (2) facilitation of students' cognitive activity, and (3) support of students' meta-cognitive learning (Praetorius & Charalambous, 2018). The theoretical foundation of the concept lies in both the application of cognitive science to educational situations and (socio)constructivist theories of learning. The primary theoretical assumptions are that to activate students cognitively, i.e., teach for understanding, the teacher must: (1) in the constructivist view, engage students in cognitive conflicts through challenging problems and questions, and (2) in the socio-constructivist view, invite students to participate in classroom discourse and communicate their ideas to develop conceptual understanding (Praetorius et al., 2018).

The notion of selection and use of tasks is further differentiated in the "offer-use-model" of teaching and learning (Helmke, 2015; Weingartner, 2021). The "offer side" refers to two aspects of teacher practice: firstly, the *objective* cognitive activation potential of tasks, which can be examined by task analysis (e.g., Neubrand et al., 2013); and secondly the *implemented* cognitive activation potential, which can be analysed by observation of classroom interactions (e.g., Kotzebue et al., 2020). The "use side" refers to the cognitive activity of students, which can be recorded by questionnaires or potentially mobile gaze tracking methods (Haataja et al., 2019). The offer-use-model highlights the reality that high levels of cognitive activation potential do not guarantee higher activity by students' actual use of the offers made by teachers.

This dissertation is primarily directed at the cognitive activation potential offered by teachers. The following sub-chapters will further explore how cognitive activation has been theorised and studied in terms of both the selection of tasks and implementation of tasks through classroom interactions, respectively.

2.2.2 Tasks as potential for cognitive activation

The theory and research on tasks as a context for student thinking has a history tracing back to the work of Walter Doyle in the early 1980s (Doyle, 1983). The underlying argument for placing a focus on tasks is that they provide a framework to define and explain cognitive activity because tasks create the context for students to think about the subject's content (Tekkumru-Kisa et al., 2020). Mathematics lessons often revolve around student work on tasks – solving problems. Pólya remarked that "solving problems can be regarded as the most characteristically human activity" (1981, p. ix). He deemed it important for teachers to differentiate between "routine" and "non-routine" problems. In Pólya's view, the non-routine problem demands some degree of creativity and originality from the student, while the routine problem does not.

In the Task Analysis Guide (TAG), another approach is taken by considering tasks in terms of the cognitive demand required to reach a solution. In TAG, mathematical tasks are divided into four categories, two of low cognitive demand and two of high cognitive demand (Stein & Smith, 1998). More demanding tasks are argued to promote analytical thinking by inviting students to justify their answers and provide reasoning. TAG is intended to analyse mathematical tasks as they appear in curricula or learning materials but not how they are represented in the classroom. The Mathematical Tasks Framework describes how mathematical tasks unfold during classroom instruction. In **Figure 1**, each rectangle symbolises a phase in the process of how tasks unfold according to the framework. TAG considers the first phase, i.e., tasks as they appear in curricula or instructional materials. The second and third phases, which consider tasks as set up by the teachers and tasks as implemented by students, are not considered by TAG. However, the second and third phases can be considered with other methods, such as classroom observations.

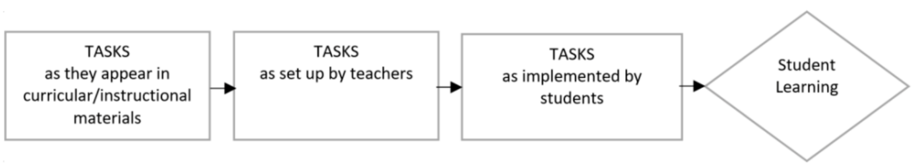


Figure 1. The Mathematical Tasks Framework (Stein & Smith, 1998).

It can be argued that the cognitive activation potential of tasks lays certain groundwork for the cognitive activity which the teacher facilitates in the classroom. It is also widely accepted that many factors contribute to student engagement and achievement, so making causal claims for different levels of cognitive demand of tasks on student learning remains problematic (Otten et al., 2017). However, analysing the cognitive demand of tasks that the teacher selects for students to engage in is useful to gain deeper insight into that specific aspect of cognitive activation. In an analysis of Icelandic textbooks in upper secondary schools, over 80% of tasks were categorised as low cognitive demand, which leaves it up to the teacher to advance and uphold the mathematical rigour of the activities should they be considered challenging (Sigurjónsson, 2014). To inquire about to what extent teachers maintain rigour, i.e., set up tasks according to the second phase of The Mathematical Tasks framework, it is necessary to observe the teacher-student interactions that take place in their classrooms.

2.2.3 Classroom interactions that facilitate cognitive activation

Cognitive activation involves not only the selection of mathematically rich and challenging tasks, but also the facilitation of students' cognitive activity during lessons. The instructional formats that teachers choose to pose tasks in are one way of how teachers choose how to facilitate cognitive activity, as well as how and to what extent they orchestrate mathematical discussion (Smith & Stein, 2011). Attending to student thinking and intervening appropriately requires the teacher to evaluate and react carefully and professionally to learners' experiences – or as Dewey stated: "It is the business of the educator to see in what direction an experience is heading" (1938, p. 38). By this view, teaching and learning occur in a social setting, where learning is a social process in the Vygotskian sense (1978), and conceptual understanding is developed through participation in classroom discourse in a socio-constructivist perspective (Praetorius et al., 2018).

Activating student thinking strongly resonates with Liljedahl's (2021) notion of a *thinking classroom*. A thinking classroom in Liljedahl's terms is a space where individuals mobilise their knowledge, think collectively and construct understanding through discussion. Among the characteristics of a thinking classroom that directly address interactions are three types of questions that students ask: (1) proximity questions – asked when the teacher is physically close; (2) stop-thinking questions – such as "is this correct?"; and (3) keep-thinking questions – asked so that students can continue their intellectual work. According to Liljedahl, only keep-thinking questions should be answered in a thinking classroom. The other types need to be acknowledged, but not answered (Liljedahl, 2018). Should a teacher answer stop-thinking questions extensively, it may reduce the challenge of the tasks and thereby lower the potential for cognitive activation. Mason and Johnston-Wilder argue that to answer such questions invites student's *accept* mode, accepting what the teacher says as truth, as opposed to *assert* mode, constructing, and exploring their own assertions and conjectures (Mason & Johnston-Wilder, 2006).

The implementation phase posits potential differences in opportunities for discourse as an element of cognitive activation. The concept of stop-thinking and keep-thinking questions in Liljedahl's thinking classroom model resemble a dichotomy of teacher moves by Furtak and Shavelson: dialogic teacher moves versus authoritative teacher moves (2009). Briefly put, authoritative teacher moves imply information transmission from teacher to students, while dialogic teacher moves promote discussions and opportunities for students to participate in knowledge construction. Opportunities for discourse, both teacher-student discourse and between students themselves, is considered an important factor of teaching quality, as teachers' responses play a crucial role in facilitating opportunities for instilling the social norm that students argue and expand on their ideas (Yackel & Cobb, 1996). Studies have found that engaging with others about their ideas enhances student learning, and student questioning, explaining, and re-explaining is associated with the growth of students' mathematical ideas (Ing et al., 2015). Classroom discourse carries weight in many observation systems, such as Schoenfeld's *Teaching for Robust Understanding* framework (Schoenfeld, 2013) and PLATO, and aims to capture different levels of social interactions in the classroom.

2.2.4 Interactions in the mathematics classroom

A couple of decades before the conceptualisation of cognitive activation, research on mathematics classrooms was already invested in the perspective of classroom interactions. This body of research was to a large extent initiated within the paradigm of social constructivism, through the traditions of symbolic interactionism or ethnomethodology (see, e.g., Krummheuer, 2012). The empirical focus of these perspectives is on the communicative exchange within the social conditions of the mathematics classroom. In raising these perspectives in mathematics education, Bauersfeld stated firmly that "teaching and learning mathematics is realised through *human interaction*" (1980, p. 35) and called for inquiry into what he called "hidden dimensions in the so-called reality of a mathematics classroom". The interactionist point of view assumes that cultural and social dimensions are not just peripheral conditions but rather intrinsic to learning mathematics (Voigt, 1995). In an outline of the core convictions of the interactionist position, Bauersfeld (1994) defined teaching as "the attempt to organise an interactive and reflexive process" and "the establishing and maintaining of a classroom culture, rather than the transmission, introduction, or even rediscovery of pregiven and objectively codified knowledge" (p. 139).

The interactionist perspective is distinct from a critical realist perspective. A critical realist perspective views language and culture as significant *mediators* in shaping the reality of mathematics teaching and learning rather than intrinsic conditions (Willig, 2013). It recognises the socially constructed nature of our understandings of reality and thus sits between a constructionist and an essentialist position (Burr, 2015). By contrast, Krummheuer (2002) has emphasised the paradigm of social constructivism in interactionist research, based on a grounded theory approach to the analysis of mathematical learning processes (Schütte et al., 2019). Such analytical work has involved both inductive and abductive inferences (Krummheuer, 2007). A critical realist approach to inductive inference may use knowledge of theories to inform coding and theme development, but the analysis is rather driven by data than theory (Terry, 2021). Typically, inductive research seeks patterns in empirical observations where the aim is to develop a theory driven by the data at hand.

Studies on interactions in mathematics classrooms have led to theorising of different *patterns* of interaction – also named types, or modes. A pattern described as occurring in an almost "incessant repetition" in classroom studies is the three-part sequence: teacher initiation, student response, and teacher evaluation (Cazden, 1988; Pimm, 1994; abbreviated IRE, or IRF as "initiation-reply-feedback"). Another pattern with strong teacher control identified particularly in mathematics teaching is known as *funnelling*, exemplified by repeated teacher questions that constantly narrow down to produce a correct answer (Bauersfeld, 1994). Voigt (1995) differentiated between *elicitation* patterns and *discussion* patterns of interaction. Elicitation patterns' main aim is the solution, with students guided to follow the teacher's way of solving step by step, thus concealing student's own competencies. Conversely, in discussion patterns, the solution is the starting point of an explanation, where student contributions are an asset to public argumentation, thus making student competencies visible. Voigt argues that in the elicitation pattern, students learn how to solve problems as expected by the teacher, while the discussion pattern creates opportunities for students to learn how to argue mathematically. These different patterns of interaction may encompass various *types* of interaction, such as teacher or student explanations, prompting with questions, or providing feedback.

Others have made a distinction between exercise-based interactions and other interactions or paradigms. Mason (2002) proposed six *modes* of interaction between student, content, and teacher, all of which were claimed to contribute to effective learning. Directly addressing teachers, the six modes proposed are:

- *Expounding*, or attracting your students into your world of experience, connections, and structure;
- *Explaining*, or entering the world of the student and working within it;
- *Exploring*, or guiding your students in fruitful directions as they sort out details and experience connections for themselves;
- *Examining*, when students validate their own developing criteria for whether they have understood, by submitting themselves for assessment;
- *Exercising*, when students are moved to rehearse techniques and to review connections between theorems, definitions and ideas;
- *Expressing*, when students are moved to express some insight (p. vii).

Some of these modes have a particular connection to cognitive activation, such as explaining, exploring, and expressing. Focusing on the last two, Mason notably distinguished between the modes of exercising and expressing, the latter referring to students expressing their understanding, e.g., by constructing examples or articulating connections between concepts. In a broader perspective, Skovsmose (2001) contrasted the *exercise paradigm* (or "traditional mathematics education") with an *investigative approach* (or "landscapes of investigation"). In the exercise paradigm, textbook work is common, with exercises formulated by an external authority, and a central premise being that only one answer is correct. Conversely, the investigative approach is a paradigm where students engage in processes of exploration and explanations that are commonly project-based and located in a "landscape" that supports investigative work.

Classroom interaction research has conceptualised interaction as patterns, modes, or paradigms. To develop a deeper empirical understanding of cognitive activation, these theoretical perspectives provide an important framing through which types of interactions can be viewed. They also offer opportunity to reflect on methodological issues of systematic observations.

2.3 Methodological background

While chapter 2.1.3 outlined literature on the use of video as a method of classroom data collection in a historical context, the way raw video data of classroom activity has generated results through the use of standardised classroom observation systems is the focus in the first part of this sub-chapter. This serves as an argument for the use of a standardised observation system in the project, whose conceptualisation of teaching quality is explained in the second part. The chapter concludes with a review of the use of surveys to capture student perceptions of some aspects of teaching, and their connection to other teaching and learning outcomes.

2.3.1 Standardised classroom observation systems

For observational analyses of teaching quality to be rigorous and consistent, a robust and reliable observation system is needed. Two critical purposes of such systems are to understand and improve teaching (Bell et al., 2019). As previously has been mentioned, teaching quality can be viewed from multiple perspectives. Wang and colleagues (2011) outlined three different perspectives of teaching quality: the cognitive resource perspective, teaching quality as performance and teaching quality as effect. The use of observation and observation systems prioritises the perspective of teaching quality as performance. The third perspective, teaching quality as effect, plays a role when it comes to student outcomes, such as achievement or perceptions. In this sub-chapter, classroom observation systems are defined and described in a historical context, with arguments for their use in this dissertation.

Classroom observation systems are comprised of scoring tools, rating quality procedures and sampling specifications used to measure teaching quality (Bell et al., 2019). Various observation systems have been employed in mathematics classroom research. Discussing qualities of many different frameworks in detail is beyond the scope of this dissertation. However, in a previously discussed analysis of twelve frameworks used in mathematics education, Praetorius & Charalambous (2018) found that all considered cognitive activation to some extent. Other teaching quality domains defined for their analysis were classroom and time management, content selection and presentation, practicing, (formative) assessment, socio-emotional support and cutting-across instructional aspects aiming to maximise student learning. Furthermore, frameworks were classified as either content-generic, mathematics-specific or hybrid, and noted that their operationalisation, purpose, approach, and theoretical underpinnings differ widely.

For a historical perspective, the pioneering TIMSS 1999 Video Study included an ambitious effort in producing an observation system with the goal of describing and investigating teaching practices in eighth-grade mathematics in a variety of countries. The system initially consisted of six dimensions of classroom practice: purpose, classroom routine, actions of participants, content, classroom talk, and classroom climate (Jacobs et al., 2003). The dimensions in subsequent analyses, after double scores of classroom videos and transcripts by certified raters, were reduced to three: purpose, classroom interaction and content activity (Givvin et al., 2005). Clarke and colleagues (2006) criticised these dimensions, both for not being independent and for being over-inclusive, to the point of being "defined in extremely simplistic terms". They suggest that the conflicting results can be resolved by "examining the nature of the interconnection of the various components of classroom practice rather than simply the frequency of their occurrence" (Clarke, 2003; Clarke et al., 2006). Furthermore, a very critical stance toward the construct of "national patterns" that was emphasised in TIMSS is evident.

The time scope and number of dimensions scored is often referred to as *grain size* and has been a contested issue for decades. Although in older conceptualisations this included a long list of teaching behaviours to score, a consensus seems to have been reached in more recent frameworks on a set of about a dozen elements (Bell et al., 2019). Another aspect of grain size is the number of scale points. One way of measuring each dimension is on a dichotomous scale, i.e., present or not present. In the field of mathematics education, such dichotomies have been subject to criticism (Blömeke et al., 2015; Clarke, 2006).

Various challenges may arise when using an observation system to analyse video-recordings of teaching. One of the challenges is when students work individually on problem sets in textbooks, a common instructional pattern in mathematics teaching Iceland (Jónsdóttir et al., 2014; Þórðardóttir & Hermannsson, 2012). Even though students may be involved in higher-order thinking processes in constructing their solutions, it is not easily captured on video with a classroom-wide lens for this type of instruction. The teacher can make a low-challenge task require higher order thinking by asking students to justify their answers and explain their solutions. The teacher can also make a high-challenge task less demanding, e.g., by telling students exactly how to solve the task or providing the answer. In such scenarios, teacher utterances in assisting individual students can weigh heavily in observation scores concerning cognitive activation. In cases where activities and assignments are not made explicit in whole-class instruction, it is important to identify and analyse the tasks that the teacher has chosen for the students to solve to preserve validity of conclusions.

White (2021) presented a validity framework for comparing teaching quality across contexts using standardised observation systems. Building on Kane's (2006) pragmatic approach to validation and an argument approach to observation protocol validity (Bell et al., 2012), White's framework highlights decisions made at all stages of a study's design and implementation that have implications for the validity of conclusions. It breaks down two key challenges: (1) operationalising and measuring teaching quality, and (2) constructing an appropriate sample of instruction in a context. Expanding on the first challenge, White highlights how the choice of observation system affects both *which* dimensions and domains of teaching quality are measured and *how* these dimensions are operationalised. Further, the rater training and monitoring process warrants careful attention as rater understanding of the rubric lens can be a source of error in observation scores (e.g., White, 2018). Decisions made in study design therefore limit its generalisability, both with regards to specific domains of teaching quality and the context to which a study aims to generalise.

2.3.2 PLATO's conceptualisation of teaching quality

Using classroom observation systems has many benefits, but the choice of a system should be argued for. One important benefit of using existing observation systems is the accumulation of knowledge through the greatly needed shared vocabulary that they provide (Klette, 2020). However, a key critique is that each system is reductionist. It is difficult to imagine an observation system capturing every single aspect of complex interactions that take place within different classrooms, so the nature of their codes is to reduce (Klette & Blikstad-Balas, 2018).

Protocol for Language Arts Teaching Observations (PLATO) is among the most widely recognised standardised classroom observation systems (Bell et al., 2019). PLATO was developed by a team of researchers at Stanford University led by Pamela Grossman and, as the name suggests, was initially intended for English language arts teaching observations (Grossman, 2015; Grossman et al., 2013). PLATO has been reliably modified by researchers for observation sampling in other subjects, including mathematics (Mahan et al., 2021). The protocol includes scoring tools in twelve dimensions (titled "elements"), each belonging to one of four overarching instructional domains: instructional scaffolding, disciplinary demand, representation and use of content, and classroom environment (see **Table 2**). Each 15-minute segment of a lesson receives a PLATO score on a scale from 1 (low) to 4 (high) for every element depending on the amount of observed evidence according to a rubric. A score at the 1-level means almost no evidence, while a score at the 4-level means strong, consistent evidence. To maintain rating quality, raters need to be certified by attending a course given by specialists in the protocol from Stanford University and passing a reliability test with a score of at least 80%.

The use of PLATO has been recommended for a variety of reasons: it resonates with existing research, its four instructional domains replicate the areas outlined in research literature as critical for student learning, it builds on a feasible and applicable observation system, it allows systematic comparison of instruction across subjects, countries, and educational settings, and it provides an opportunity to test for possible cultural biases embedded in the PLATO instrument (Klette et al., 2017). Furthermore, PLATO's 4-point scale eliminates the need to dichotomise individual teaching quality elements. Such dichotomisation has been subject to criticism in classroom research as it does not capture the diversity of teaching practice (Blömeke et al., 2015; Clarke, 2006; Kyriakides et al., 2013).

The instructional domains in PLATO are seen in **Table 2**. As previously identified (Bell et al., 2019), cognitive activation is primarily reflected in PLATO elements in the instructional domain entitled "disciplinary demand" (highlighted in **Table 2**). This domain includes three elements: "intellectual challenge" (IC), "classroom discourse" (CD) and "text-based instruction". Due to the nature of mathematics as a subject, text-based instruction is not scored in the mathematics-adapted version of PLATO. The use of the two relevant elements for the aims of this research project, IC and CD, is explained in detail in chapter 3.5.1.

Table 2. All PLATO elements and their respective instructional domains.

PLATO instructional domains	PLATO elements
Instructional Scaffolding	Modelling Strategy Use and Instruction Feedback Accommodations for Language Learning
Disciplinary Demand	Classroom Discourse (CD) Intellectual Challenge (IC) Text-Based Instruction
Representation and Use of Content	Representation of Content Purpose Connections to Prior Knowledge
Classroom Environment	Behaviour Management Time Management

2.3.3 Student perceptions of teaching quality

Although much can be learned from evaluating tasks used in classrooms and observing the lessons where they are implemented, a limitation to these approaches is their time intensity. Out of the myriad of lessons taught by teachers, it is only feasible for researchers to observe a few. It has been acknowledged that teaching quality is a complex and multidimensional phenomenon whose study requires a variety of complementary strategies (Croninger et al., 2012). A less time-consuming approach is considering the students' perceptions. Students gather experiences in the classroom every day and can provide important information about teaching that may not be captured by classroom observations.

Student perceptions of teaching have been an object of study since at least the 1970s and are widely considered an important element of classroom research (den Brok et al., 2006). In recent years they have received increased attention by researchers, not least due to the implementation of a student survey in the Measures of Effective Teaching (MET) study (Ferguson, 2010). The Tripod survey used in the MET study has been developed and modified since 2001 and administered to over one million students since then. Addressing the concern of whether student surveys can measure teaching quality reliably, Ferguson claims that the results of the MET study should put such doubts to rest and anticipated a growing consensus among researchers in implementing student perception surveys in measuring teaching effectiveness (Ferguson, 2012). Strong correlations between the seven Tripod dimensions have suggested that the framework can be reduced to two dimensions at the between-classrooms level: classroom management and instructional support (Schweig, 2014; Wallace et al., 2016). Further research has indicated that the validity and reliability of student perception scales can be high even at the primary level, although further research is needed on correlations between different scales (van der Scheer et al., 2019).

The connection between student perceptions of teaching and other teaching quality measures has been investigated with somewhat mixed results. In the GTI study (formerly TALIS), most students perceived high cognitive engagement in lessons despite most observation scores being low for cognitive engagement, showing a discrepancy between the two measures (OECD, 2020). Data from the MET study were used to compare student perceptions as measured by the Tripod survey to observation scores in middle school mathematics lessons. Among the three instructional domains observed, a low correlation was found between general student perception to emotional support and instructional support. No correlation was found between student perception and classroom management (Wallace et al., 2016). Another study, drawing on the same dataset, examined correlations between the proposed seven dimensions of Tripod and both two observation systems and value-added measures. The correlations between Tripod dimensions and observation system dimensions were mostly low to medium. Notably, the Tripod dimension "challenge", concerning rigour and persistence, was positively related to value-added measures, but score variability within the dimension was negatively related to value-added measures (Sandilos et al., 2019). Furthermore, models have indicated a higher expected correlation between the two approaches with a higher number of observations (van der Lans, 2018). In conclusion, these mixed results show a somewhat unclear connection between student perceptions and observation scores and explain why caution has been advised in using student surveys for high-stakes teacher evaluations (Kuhfeld, 2017; Phillips et al., 2021).

2.4 Summary and research questions

In this chapter, a review of literature has been presented in three main strands of research. First, the context of the study was described by summarising previous research on mathematics teaching within the Icelandic and Nordic school contexts, as well as contextualising the research project with reference to previous video-based studies in mathematics education. Findings on mathematics teaching in Iceland have shown that there is room for improvement, with lesson time dominated by procedural individual work, poor outcomes in international measurement, and disappointing results of prospective teachers' knowledge. Studies on mathematics teaching in the Nordic countries have interestingly shown rather similar results, with a strong emphasis on procedural fluency and individual work, which suggests cognitive activation at a low level. Finally, previous large-scale studies based on video-recordings of mathematics lessons were discussed, providing historical context of this research method. Some findings include that video enables in-depth study of complex processes, that analysis of video data is highly time-consuming, and that claims of "national patterns" of teaching are particularly problematic.

Secondly, theoretical perspectives of teaching quality, particularly cognitive activation, were presented in context with literature on interactions in mathematics classrooms. Cognitive activation refers to what extent a teacher addresses the educational goal of student understanding. In a constructivist perspective, this may encompass the cognitive potential of tasks, but in a socio-constructivist perspective also the teachers' facilitation of students' cognitive activity and metacognitive learning. The socio-constructivist perspective was in the forefront in interactionist research in decades prior to the conceptualisation of cognitive activation. This body of research proposed different patterns, modes, or paradigms of interactions between teachers and students, which may also be viewed with a critical realist perspective.

Thirdly, literature was reviewed in relation to methodology. Standardised classroom observation systems were discussed, a definition offered and issues relating to their design, validity and reliability discussed. Student perceptions of teaching are widely employed in classroom research and considered unique insights for researchers, but results on their connection to other outcomes such as observation scores have been mixed.

As outlined in chapter 1.3, the overarching aim of the research project was to develop a deeper understanding of the teaching quality dimension of cognitive activation in mathematics in Iceland and in a Nordic context. This chapter concludes with the research questions proposed and addressed in each research paper, and their connection to the aims:

I. How can the potential for cognitive activation in lower secondary mathematics in Iceland be described? (Aims 1 & 2)

II. What is the nature of the connection between classroom observations and student ratings as measures of cognitive activation? (Aims 3 & 4)

III. What characterises teacher-student interactions in lower secondary mathematics lessons considered outstanding in cognitive activation in a Nordic context? (Aims 5 & 6)

3 Methods

In this chapter, the methods chosen to reach the aims of the research project are explained. First, an explanation is offered of how the study is situated as part of a larger research initiative. Next, a detailed account is made of the research design, participants, collection of data and the various analyses made to these data. The chapter concludes with a discussion of ethical considerations and limitations of the project.

3.1 Situating the research project within QUINT

This dissertation is based on my cooperation with the Quality in Nordic Teaching (QUINT) research centre as a PhD fellow within the LISA Nordic study. QUINT's vision is to systematically research teaching quality in the Nordic countries using both classroom video-recordings and student surveys. One of the QUINT ambitions is the LISA Nordic study, which aimed to build a Nordic classroom video database that can be analysed from different perspectives. The dataset can be accessed through a specific procedure (Klette, 2022). To build the database, lessons were video-recorded in language arts, mathematics, and social science in ten schools from each country: Iceland, Sweden, Norway, Denmark, and Finland. Three to four consecutive lessons in each subject were video-recorded. A student survey was also administered to each student group whose lessons were video-recorded (students aged 13-14; grade 8 in Iceland). As a QUINT PhD fellow, I received training in video data collection and participated in the collection of video and student survey data in two schools in Iceland in spring 2019. Other Iceland data were collected by fellow QUINT researchers from University of Iceland and University of Akureyri. Similar data were collected by QUINT researchers from partner universities within each Nordic country. **Figure 2** gives an overview of the analogous data collection for the LISA Nordic study in each country, from collection of video and survey data to observation scores in a central QUINT database. The data collection process is explained in more detail in chapter 3.4.

As this dissertation draws on the mathematics part of the LISA Nordic dataset, only that part of the dataset is reported on. In sum, LISA Nordic includes over 150 video-recorded mathematics lessons across the Nordic countries. This dissertation draws mostly from the Iceland data, but also on specific lessons from Sweden, Norway, and Denmark. Initially, the intention was to include Finland data as well, but due to delays in data collection resulting from the effects of the COVID-19 pandemic, it became infeasible.

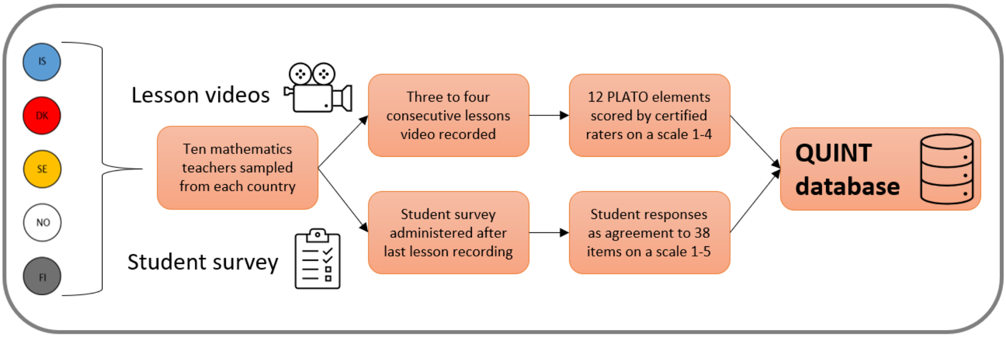


Figure 2. QUINT's collection of data in each country for the LISA Nordic study.

By drawing from a larger dataset for the doctoral research project, the analyses are largely *secondary data analyses*, i.e., handling of data collected by others (Smith, 2011). This means that, apart from two classrooms in Iceland, I have no first-hand personal experience of being in those classrooms. Secondary data analyses can provide distinctive opportunities, such as in-depth inquiry into uniquely sampled data – but also challenges, such as having adequate contextual information, fit of research aims to the data available, and ethical issues (Andersson & Sørvik, 2013). A great amount of contextual information was gathered through data logs and the collaboration of the QUINT Iceland group which collected and analysed the Iceland data. Good cooperation with the colleagues who translated and transcribed the Nordic lessons also provided me with very useful contextual information. A certain gap between what was observable in the videos regarding the cognitive activation potential of tasks was addressed with additional data generated on mathematical tasks (reported in Paper I). Otherwise, the research aims fit quite well with the data available. Ethical issues are specifically discussed in chapter 3.6.

One of the decisions made by QUINT was that the entire video dataset was to be systematically analysed using the same observation system, PLATO (described in chapter 2.3.2). In fall 2019, I attended a PLATO course given by specialists in the system from Stanford University. The Iceland mathematics lesson data were rated by me and two other certified PLATO raters in spring 2020, of which I rated around a third. To reach the aims of this research project focusing on cognitive activation, two specific elements of PLATO were relevant: Intellectual Challenge (IC) and Classroom Discourse (CD). They are explained in more detail in chapter 3.5.1.

As this doctoral research project is partly qualitative, it is necessary to reflect on my perspective and positionality. My background informs my somewhat pragmatist positioning – an aspiration to support developments toward improved mathematics teaching practices. I entered doctoral studies as a mathematics educator myself, with a teaching certificate, specialisation in mathematics, and teaching experience from both university and lower secondary school levels. During my studies, I have found that my perspective on teaching and learning may be best described by radical constructivism (Glasersfeld, 1995). This entails that learning is knowledge constructed by individuals, i.e., "knowledge is in the heads of persons" – although I do see the social aspect of learning as important as well. This aligns well with the theoretical grounding of cognitive activation in constructivist and socio-constructivist views of learning and teaching practice (Praetorius et al., 2018).

To analyse the data on teacher-student interactions, I take the perspective of critical realism. In the realist sense, this fundamentally means that there is a single truth and one reality – but critically, it is *situated* and *mediated* by a social construction through language and culture yet *not defined* by it. I argue that this choice is consistent with my pragmatic approach, as I assert that the single "real world" may be uniquely interpreted by individuals. As Morgan (2007) points out, this suggests a "reflexive" orientation which is precisely the form of thematic analysis I have chosen to apply to the teacher-student interactions. Even though the content analysis of the lessons portrays a timeline of time spent in each instructional format to give an overview of each lesson's format and design, I argue that the heart of the analysis is more aligned with a "big Q" qualitative framework. The focus is on what the participating teachers do and say during lessons – it does not focus on their being, who they are and what they feel, but rather on how their actions shape the classroom reality. This assumes a *constructionist* perspective of language – use of language is an active agent in creating meaning (Braun & Clarke, 2022). This means that the focus of the analysis is on the qualitative nature of interactions and instructional formats as social objects (Leung & Chung, 2019). Exact counts of minutes or types of interactions are therefore not considered of importance, but rather their meaning and relation to the outcome; a lesson considered cognitively activating. Numerically, it suffices to report that lesson length varied from 45 to 70 minutes, and the coded teacher-student interactions in the lessons totalled to around 550 interactions.

A vital aspect of reflexivity is to consider the researcher's positionality toward the research topic and the participants, as an "insider" or an "outsider" (Braun & Clarke, 2022). Because I am educated as a mathematics teacher and have personally taught lower secondary mathematics in Iceland, including grade 8, one could say that I am, in a sense, an insider to the topic and participants. I have completed a similar teacher education program as many of the participating teachers and been assigned to teach the same content as them – and in the case of the Iceland teachers I have taught according to the same curriculum as them. Therefore, me and the participants have a certain shared experience. In many ways, this is an advantage. I have a deep appreciation and understanding of the challenges they face in their mathematics teaching. In contrast, I am not currently employed as a lower secondary school teacher, and further, I have never worked as a teacher outside of Iceland. During the analysis, I was employed as a doctoral graduate student, with a point of view as an educational researcher rather than a practicing lower secondary teacher. Therefore, I may view the video-recorded lessons with a different lens than the teachers would. I also had very limited knowledge about the students they were teaching. In that sense, one could also consider me an outsider – although I do share a certain background and experience with the participating teachers as an insider. This mixed position is important to state, as it has a role to play in the way I have interpreted the data and therefore the results and conclusions I have made.

From a classroom interaction perspective, among seven categories of research approaches that have been identified (Rex et al., 2006), this qualitative part of the dissertation perhaps best aligns with a "constructivist cognitivist perspective". This includes an interpretive approach of classroom interactional practices that assumes teachers strive to create a learning environment to forward learning by active student engagement. Rather than a purely inductive orientation to the analysis of teacher-student interactions, I argue that *abductive* reasoning was also applied. Abductive inference, sometimes called *inference to the best explanation*, can lead to explanations that are plausible but not verifiable. In research on complex social settings, such as classroom teaching, one can argue that abductive reasoning is a plausible approach to create novel insights aimed at practical implications (Timmermans & Tavory, 2012). Thus, it aligns well with my pragmatic positioning. A detailed description of the qualitative analysis is in chapter 3.5.4.

3.2 Research design

To reach the overarching aim of the research project and to develop a deeper understanding of cognitive activation in mathematics teaching in Iceland and in a Nordic context, specific research approaches were required. QUINT's design and data collection for the LISA Nordic study in mathematics (in **Figure 2**) allowed for designing a research project with this overarching aim. **Figure 3** charts how the data from the QUINT database were used to design the doctoral research project on which this dissertation reports. The green part on the right side of the figure highlights decisions made personally as a researcher as separate from (though aligning with) the QUINT ambition. The roman numerals signify the research question addressed for each research output, i.e., Papers I, II, and III. Essentially, the PLATO coding of the lessons involves quantifying data that are qualitatively rich in nature. However, the research design allowed for certain qualitative richness to be preserved, though differently between the research questions addressed in each paper. By doing this, a mixed methods approach was taken in reaching the doctoral research project's aims (Buchholtz, 2019; Morse, 2003). The following list explains how each research question addressed specific sub-aims and how that may be described from a mixed methods perspective. This is further summarised in **Table 3**, outlining the data sources, units of analyses and research methods for each one in connection to the research aims 1-6.

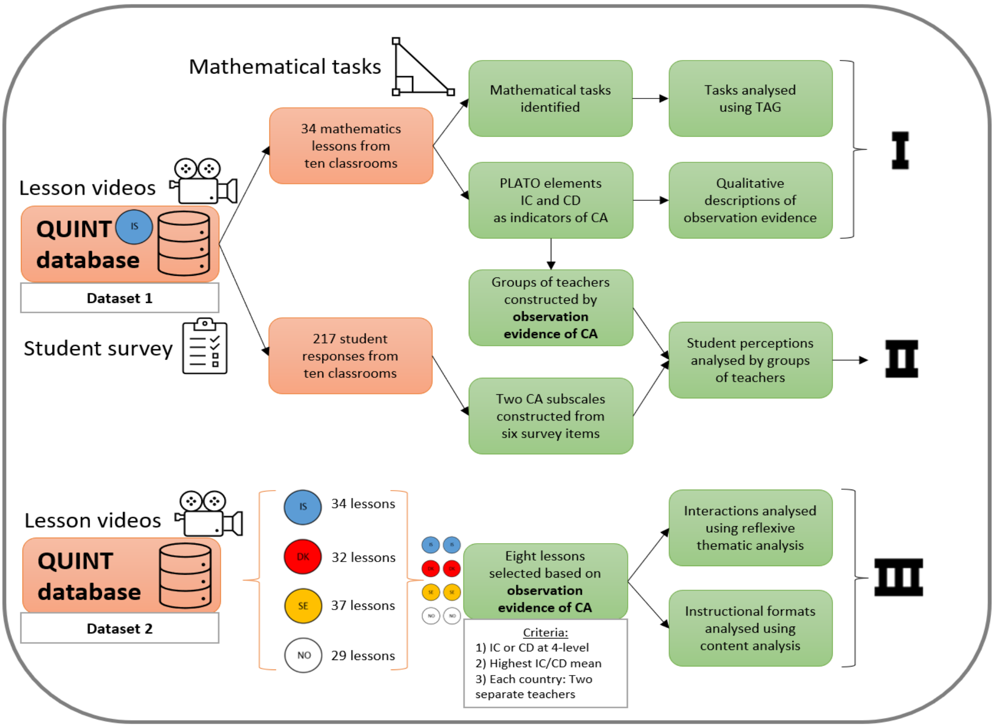


Figure 3. Design of the doctoral research project, with roman numerals signifying the research questions I, II, and III, addressed in each research paper.

- Research question I (addressed in Paper I) addressed the aims 1 and 2: to assess the cognitive activation potential in both tasks and in mathematics teaching in lower secondary classrooms in Iceland. The data used for this study were classroom video data from ten Icelandic mathematics classrooms and the tasks that were identified in the recorded lessons. While the data were qualitative in nature, the analytical frameworks worked to quantify and categorise the data. These mixed methods generated results that were numerical (quantitative) but were accompanied by qualitative descriptions to further illuminate and provide argumentation for the results – what may be called a "small q" qualitative design (Braun & Clarke, 2022). The design may be described as [qual + QUANT (PLATO + TAG)].
- Research question II (addressed in Paper II) addressed the aims 3 and 4: to examine students' perceptions of cognitive activation and its connection to the observed level of cognitive activation in mathematics teaching. The data used for this study were the quantified codes of the same classroom video data as in Paper I, and the student survey data from the ten Icelandic mathematics classrooms. As a result, the second study was predominantly quantitative in nature, using a comparison between the coded classroom level data and the student level survey data. The design may be described as [qual -> QUANT (PLATO + Tripod)].

Table 3. Summary of data sources, units of analyses and research methods for reaching the aims of the doctoral research project.

Research aims	Research question	Data	Unit of analyses	Research methods
Aim 1: To assess the cognitive activation potential of mathematical tasks in lower secondary classrooms in Iceland. *Aim 2: To assess the level of cognitive activation in mathematics teaching in lower secondary classrooms in Iceland.*	*I. How can the potential for cognitive activation in lower secondary mathematics in Iceland be described?* (Paper I)	Video-recorded lessons Mathe-matical tasks	15-minute segments (N = 88) of 34 mathematics lessons from Iceland Individual tasks (N = 144)	Descriptive Mixed Systematic video observations Task analysis
Aim 3: To examine student perceptions of cognitive activation. *Aim 4: To examine the connection between observed level of cognitive activation and student perceptions.*	*II. What is the nature of the connection between classroom observations and student ratings as measures of cognitive activation?* (Paper II)	Video-recorded lessons Student survey responses	Teaching in classrooms (N = 10) [Teachers: observation scores; Students: survey (N = 217)]	Comparative aspect Quantitative Systematic video observations Student survey
Aim 5: To enrich empirical understandings of instructional formats in Nordic mathematics lessons considered cognitively activating. *Aim 6: To enrich empirical understandings of teacher-student interactions in Nordic mathematics lessons considered cognitively activating.*	*III. What characterises teacher-student interactions in lower secondary mathematics lessons considered outstanding in cognitive activation in a Nordic context?* (Paper III)	Video-recorded lessons Lesson transcripts	Interaction episodes from purposefully chosen lessons	Descriptive Qualitative Reflexive thematic analysis Content analysis

- Research question III (addressed in Paper III) addressed the aims 5 and 6: to enrich empirical understandings of instructional formats and teacher-student interactions in Nordic mathematics classrooms. Specific video-recorded lessons from Iceland, Sweden, Norway, and Denmark were chosen with criteria based on quantified codes but analysed qualitatively with reflexive thematic analysis (Braun & Clarke, 2022) and content analysis. The criteria for a lesson to be considered, in order, was: 1) The lesson had an IC or CD segment score at the 4-level, 2) The lesson's mean score for IC and CD, 3) Two lessons from two separate teachers selected from each country. The results were built on themes that were developed based on a thematic analysis of interactions between teachers and students concerning mathematical content in the recorded lessons, and content analysis of instructional formats. The design may be described as [quant -> QUAL].

A mixed methods approach to research design presents both challenges and benefits. One of the main challenges is that it requires extensive data collection and careful sampling decisions. Combining quantitative and qualitative methods also creates a challenge in writing articles and reports. This is because it requires one to bring different research paradigms into alignment in arguing for how they are an adequate method for developing an understanding of the research topic. I argue that in the case of this project the benefits have justified overcoming these challenges. Mixed methods enable examination of a complex research topic such as cognitive activation as a dimension of teaching quality by using data comprehensively and from different perspectives. As McMillan & Schumacher also point out, multiple approaches compensate for disadvantages in using a single method, which in turn allows for investigation of different types of questions in a single study (McMillan & Schumacher, 2014).

3.3 Participants

As stated in in the previous sub-chapter, the participants came from two datasets: Dataset 1 being participants in the Iceland part of the data for research questions I and II, and Dataset 2 the Nordic participant teachers whose lessons were selected to address research question III.

Dataset 1 included participating teachers and students from grade 8 mathematics classrooms from ten compulsory schools in Iceland. The ten schools were purposefully sampled to achieve variety in several school variables, such as school location, size, organisation and social background. Half of the schools were from the greater Reykjavík urban area, and half from more rural areas. The school size varied from schools with less than 300 students (considered small in a Nordic context) to schools with over 500 students. A couple of schools were open plan schools, and one had team teaching as well as open plan. One school had a 20% proportion of students with immigrant backgrounds, which is high in an Icelandic context.

Students in grade 8 in Iceland are 13-14 years of age. They participated by being present in the video-recorded lessons and by responding to the survey. **Table 4** presents information about the participating teachers in Iceland, i.e., their teaching experience, gender, age, and whether they are specialised in mathematics teaching or not.

Table 4. Teaching experience, gender, age, and mathematics teaching specialisation of the participating teachers in Iceland.

Teacher[d]	Teaching experience (years)	Gender	Age	Specialisation in mathematics teaching[e]
T_1	14	Female	30-39	✓
T_2	10	Female	30-39	✓
T_3	11	Male	30-39	✓
T_4	16	Female	40-49	
T_5	16	Female	40-49	✓
T_6	33	Male	60+	
T_7	1	Female	30-39	
T_8	4	Female	30-39	✓
T_9	1	Female	30-39	
T_{10}	28	Male	50-59	

Dataset 2 included the Nordic participants in the specifically selected lessons from the common QUINT dataset (Klette, 2022). Dataset 2 included eight mathematics teachers, of which two teachers were from the Iceland grade 8 sample in dataset 1. The remaining six were mathematics teachers of students in equivalent grade levels from Sweden, Norway, and Denmark.

Table 5 shows the same information as in **Table 4** for the Nordic teachers. They are given pseudonyms where the first letter of the name represents the country in which the lesson was recorded (Í for Iceland, S for Sweden, N for Norway, and D for Denmark). It is noteworthy to observe that all the teachers in Dataset 2 were specialised in mathematics teaching.

[d] The numbering is ordered based on findings described in chapter 4.1.1.

[e] At least 31-60 ECTS credits in mathematics or mathematics education, which is equivalent to approximately one year of university studies. The same criterion applies in **Table 5**.

Table 5. Teaching experience, gender, age, and mathematics teaching specialisation of the teacher participants in the selected Nordic lessons.

Teacher	Teaching experience (years)	Gender	Age	Specialisation in mathematics teaching
Ída	10	Female	30-39	✓
Íris	14	Female	30-39	✓
Sabrine	8	Female	20-29	✓
Sandra	19	Female	50-59	✓
Nadia	2	Female	30-39	✓
Nils	11	Male	40-49	✓
Daniel	4	Male	30-39	✓
Doris	21	Female	40-49	✓

The process of contacting participants was as follows: First, the principal was contacted via e-mail with an invitation and detailed information about the study. The school leaders approached the teachers. If the teachers responded positively, a member from the Icelandic QUINT research team met with them. If they were ready to participate, an informed consent was given by teachers, student, and legal guardians (see Appendix E). Only when this was in place did the data collection take place. This process was analogous in all Nordic countries participating in the QUINT data collection.

3.4 Data collection

The dissertation relied on three primary data sources: video-recordings of lessons, mathematical tasks identified in the lessons, and a student survey administered to students in the recorded lessons. The data collection in Iceland took place from March through May 2019. The data were collected by a group of researchers in the QUINT research project from both University of Iceland and University of Akureyri. Two researchers went together to collect data in each school. I took part in collecting data in two schools. Equivalent data collection took place in the other Nordic countries by researchers at partner universities in QUINT. Some additional data collected, such as photographs and data logs, served a contextual purpose for this dissertation.

3.4.1 Video-recordings of lessons

Three to four consecutive mathematics lessons were video-recorded in each school (n=34 total lessons in Iceland; double-slot lessons count as two). The lessons were video-recorded with two cameras and two microphones, connected through a mixer and synchronised using specialised software. Typically, one camera was placed at the back of the classroom and aimed at a whiteboard, while the other was placed in the front and aimed in the opposite direction to capture the entire class (see **Figure 4**–inspired by van Bommel's figure in Tengberg et al., 2021). One microphone was attached to the teacher, with the other microphone typically hanging above the class to capture other dialogue between students in the classroom.

The collection of video data aimed to capture lessons without interfering or giving teachers any instructions other than to teach as they normally would. The approach strongly resembles Clarke and Chan's (2019) metaphor of "video as window" into classroom reality, where the researcher acts as a neutral witness of the social interactions with minimal intrusion. They argue that the danger of this metaphoric role of video is the proposed "neutrality" of the researcher. Indeed, however clear the view through the "window" of video into the classroom is, what the researcher sees is filtered through both the placement and angle of the video and through the frameworks that the researcher uses to generate and analyse data. As **Figure 4** shows, in the QUINT data collection design both cameras were stationary as well as the ceiling microphone that aimed to capture student talk. The teacher's microphone moved with the teacher. As the focus of the dissertation is on the actions of the teacher, the teacher's microphone and the (typically) "whiteboard" camera were primarily relied on.

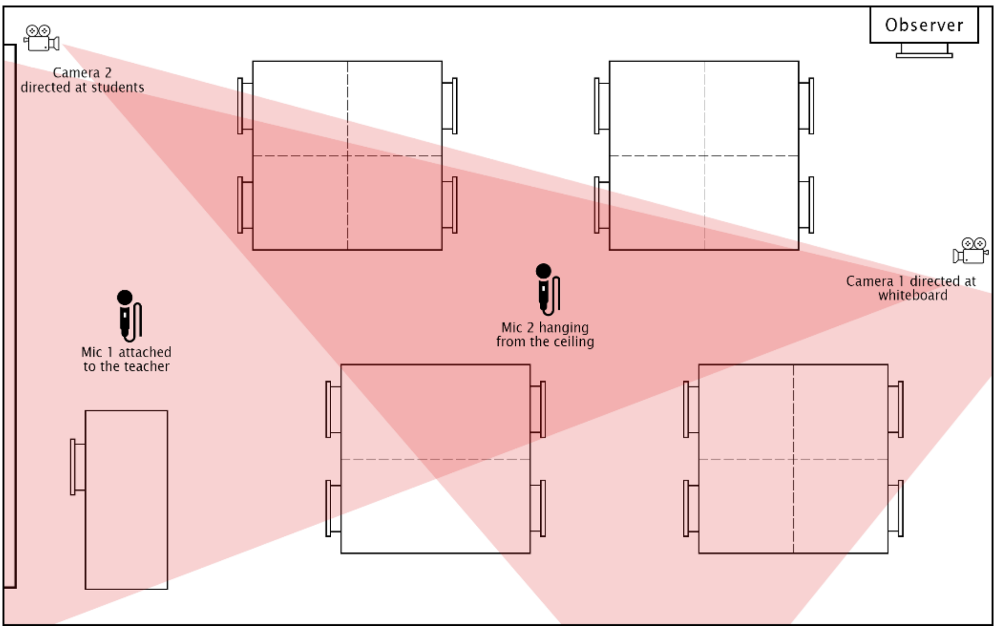


Figure 4. Camera and microphone set-up for data collection.

3.4.2 Tasks and contextual data

Some additional data were also collected in the QUINT design. These included photographs of classroom artefacts, such as whiteboard writings and student work, lesson plans from teachers, and teacher interviews intended to collect data on their background and to discuss the typicality of the recorded lessons. For my purposes, these data served to provide context and were not specifically used for analysis. However, the lesson plans sometimes proved useful in identifying the selected tasks in the lessons, and the interviews enhanced the validity of the video analysis as none of the teachers reported the lessons being atypical for their teaching. Each video-recorded lesson was accompanied by data logs, with both directly observed data from the site and derived from the teacher interviews. These data logs included metadata on the lessons, such as content area and teaching methods, as well as data on the teachers' background.

The mathematical task data were collected in two ways: by identifying numbered exercises used as observed from the video-recordings or contextual data to find the tasks in the relevant textbooks, or by asking teachers to send us more detailed lesson plans or specific tasks that were worked on. The second way applied specifically if selected tasks were teacher-made or otherwise not readily available from official textbooks or other accessible resources. Only tasks that were observed to be used in the recorded lessons were assembled and listed for analysis.

3.4.3 The Tripod student survey

A student survey was also part of the data collection. At the end of the last recorded lesson in each classroom, students (n=217 respondents in mathematics in Iceland) were asked to respond to the Tripod student survey, initially developed by Ferguson (2010). The survey is based on the 7 C's framework of effective teaching, referring to the seven components of teaching practice measured in the survey: care, confer, captivate, clarify, consolidate, challenge and classroom management (sometimes labelled "control"). The survey was used in the Measures of Effective Teaching (MET) study to measure student perceptions of teaching effectiveness and was claimed to successfully measure teaching quality and predict student achievement (Ferguson, 2012). The survey's purpose is to assess to what extent the students experience the classroom environment as engaging, demanding and supportive for their intellectual growth. The students stated the frequency of different actions or activities in the classroom that indicated their perceptions of teaching. For data collection organised by QUINT, the survey was slightly modified by adding two questions pertaining to the "care" component, making the total number of items sum up to 38 (the full list of survey items is in Appendix C). The students respond on a five-point ordinal scale: never, rarely, sometimes, often, always. The translated and slightly enhanced survey was piloted in Norway before the QUINT data collection (Klette et al., 2017). **Table 6** shows descriptive data on the participating schools and the number of student responses from the video-recorded mathematics classroom in each school in Iceland. All students present in the lessons in Iceland agreed to respond to the survey.

Table 6. Number of mathematics lessons recorded by schools, total length of recordings in each classroom, and number of student survey responses.

Teacher	Number of lessons	Total video length (minutes)	Total survey responses
T_1	4	112	26
T_2	3	165	9
T_3	2	76	42
T_4	4	145	18
T_5	3	154	27
T_6	4	127	11
T_7	3	147	15
T_8	3	112	28
T_9	4	146	15
T_{10}	4	154	26
Total	34	1338	217

3.4.4 Selection of outstanding lessons

After a broad view was taken across 34 mathematics lessons collected in Iceland, an in-depth look at specific lessons from each country that stood out in cognitive activation was deemed to be a more fruitful path toward empirically rich findings for aims 5 and 6 rather than an even broader view across countries. Thus, instead of trying to quantitatively compare teaching using observation scores across the gargantuan Nordic dataset, specific lessons were selected with high segment scores in cognitive activation (IC and CD). Even if the entire dataset were to be considered, to draw conclusions about teaching in each entire country would be riddled with difficulties (Xu & Clarke, 2019). To explore research question III, there was a need to construct selection criteria to identify and select the lessons in the Nordic database with the strongest evidence of cognitive activation to explore research question III.

The criteria constructed to select outstanding lessons in cognitive activation were threefold. The first criterion was that a lesson contained a segment score at the 4-level in either IC or CD. This filtered out a vast majority of lessons in the database. The second criterion was a comparison of the lesson's mean score value for both IC and CD. The third criterion was to select two lessons from two separate teachers in each country. In the case of Denmark, there were no lessons with scores at the 4-level in IC or CD. The first criterion was then modified for a 3-level segment score before considering the mean score. **Figure 5** shows a scatterplot of PLATO scores in IC or CD in all mathematics lessons from these countries in the LISA Nordic database. Each dot represents a lesson's mean score in IC and CD with the selected lessons highlighted.

Of eight specifically selected lessons from the Nordic database, six were from other countries than Iceland, i.e., Sweden, Norway, and Denmark. Access was granted to the video database through collaboration with the data manager at the University of Oslo. A secure remote connection was established to a computer based in their video lab. The chosen lessons were translated and transcribed into English by two collaborators in the QUINT project. These research colleagues also provided contextual information as well as confirming mutual understandings of instances of ambiguity, either in the transcripts or the video data itself, furthering the credibility of the results. The Iceland mathematics lessons were transcribed by me and a research assistant. All personal identifiable information was either deleted or altered in the transcripts before analysis.

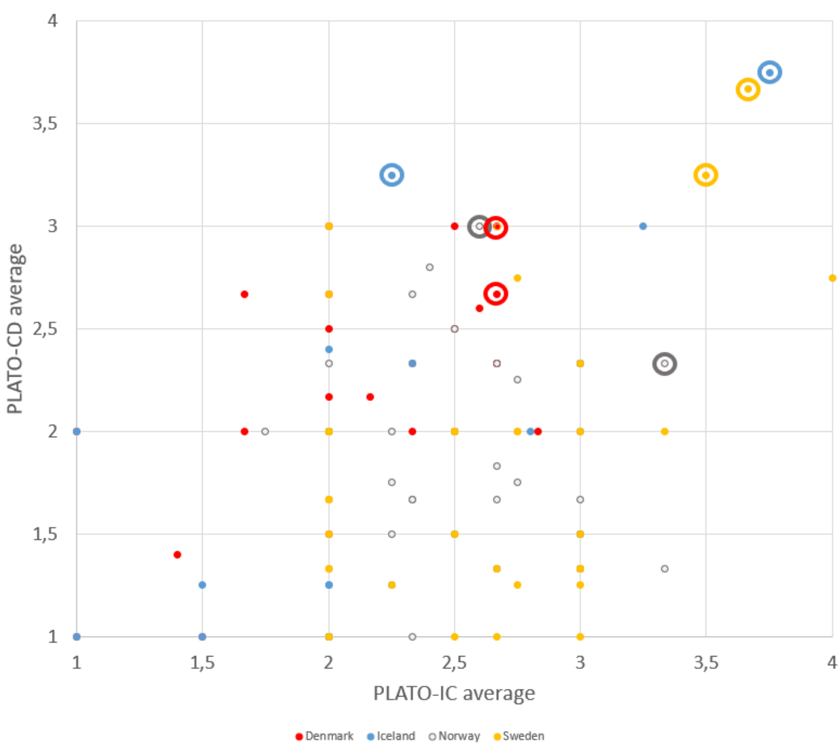


Figure 5. Mean scores in IC and CD across all LISA Nordic mathematics lessons with selected lessons highlighted.

3.5 Data analysis

Different frameworks were applied to each of the three primary data sources: classroom video-recordings, mathematical tasks and the student survey. All lesson recordings were analysed with PLATO and the results in the IC and CD dimensions were relevant for all research questions. Mathematical tasks were analysed using TAG for aims 1 and 2. Specific items from the Tripod student survey were analysed statistically in comparison with the PLATO scores for aims 3 and 4. For aims 5 and 6, the selected lessons with outstanding PLATO scores were analysed qualitatively, using content analysis and reflexive thematic analysis.

3.5.1 Cognitive activation through classroom video observation (PLATO)

PLATO was presented generally in chapter 2.3.2, but its specific use for analysis for this dissertation is discussed here. Coding lesson videos with PLATO means raters use observation evidence to apply a score of 1 (low) to 4 (high) to each 15-minute segment of a lesson for each of PLATO's 12 dimensions. The elements of PLATO that relate to cognitive activation and are therefore in focus in this dissertation are Intellectual Challenge (IC) and Classroom Discourse (CD), which belong to the PLATO instructional domain "disciplinary demand".

The choice to frame the research project in terms of "cognitive activation" instead of "disciplinary demand" is based on two reasons. First, a shared vocabulary has been greatly needed in the field of research on teaching (Grossman & McDonald, 2008; Klette et al., 2017). To use cognitive activation means framing the research with a more generally known concept and dimension of teaching quality than a concept and instructional domain mostly known within a specific observation system. The second reason is connected to the conceptualisation of the first. Recent suggestions for a common structure for comparing frameworks of teaching quality (presented in chapter 2.2.1) defined cognitive activation trough a teacher's:

1. a) selection of challenging tasks that respond to the students' cognitive level, and b) use of mathematically rich practices
2. facilitation of students' cognitive activity
3. support of students' meta-cognitive learning from cognitively activating tasks (Praetorius & Charalambous, 2018)

The second reason is that, based on this definition, it can be justified to use the IC and CD elements to analyse cognitive activation from lesson observations.

In PLATO, the IC and CD elements address all three points from the definition above, except there is no explicit measure of meta-cognitive learning support. IC focuses on the intellectual rigour of activities that students engage in. The score can be advanced or degraded by one point based on the nature of teacher questions and comments in relation to the challenge of the task as initially presented. High-level IC involves students engaged in analytic or inferential thinking, while low-level IC refers to activities where students mostly engage in recall or rote thinking. In this way, IC addresses the selection and use of tasks and the richness and rigour of facilitated cognitive activity. The CD element focuses on both the opportunities students have for extended mathematics-related talk (with the teacher or other students) and to what extent the teacher and students pick up on, build on, and clarify each other's ideas. High-level CD has students engaged in focused discussions where the teacher and students build on each other's contributions and ask for ideas to be clarified and specified. Low-level CD usually means either that the teacher does most of the talking or that the talk is disconnected, with the teacher and students not building on previous responses (Grossman, 2015). Therefore, CD addresses the facilitation of cognitive activity and use of mathematically rich practices, while IC has a stronger focus on the tasks. In **Table 7**, a shortened version of the rubric is portrayed for each level in both IC and CD. Appendix A shows the full rubrics for IC and CD (the full manual with detailed scoring procedures is only available upon request from PLATO's creators).

Table 7. Scoring rubric (shortened) for PLATO elements IC and CD (Grossman, 2019).

	Intellectual Challenge (IC)	**Classroom Discourse (CD)**
	Teacher provides activities or assignments that are...	Teacher or students...
1-level	almost entirely rote or recall	rarely if ever respond to students' ideas about mathematical content. Few to no opportunities for mathematics related student talk.
2-level	largely rote or recall, a portion of the segment promotes analysis, interpretation, inferencing, or idea generation	respond briefly to student ideas. Talk is tightly teacher-directed. Occasional opportunities for mathematics related student talk.
3-level	a mix: most promote analysis, interpretation, inferencing, or idea generation; a few are focused on recall or rote tasks	show multiple instances where student ideas are specifically addressed. There are opportunities for mathematics related student talk but may be substantial teacher direction.
4-level	rigorous and largely promote sophisticated or high-level analytic and inferential thinking, including synthesising and evaluating information and/or justifying or defending their answers or positions	consistently engage in high-level uptake of student ideas. Most students participate by speaking or actively listening and students respond to each other. Open-ended questions, a clear focus and on-track conversation.

The lessons recorded by QUINT were scored by certified raters in each Nordic country. I became a certified PLATO rater in October 2019 after completing a one-week course given by specialists in the system and passing the 80% reliability test. Inter-rater reliability was periodically checked with double coding. The lessons from each school were divided between two researchers, so that researcher A coded lessons 1-2 and researcher B coded lessons 3-4. Both researchers coded the first segment of lessons 1 and 3 and met to compare their scores. If inconsistencies were found, the justification of the scores were discussed to find a mutually agreed upon score that aligned with the PLATO protocol. Joint video workshops were also organised by the QUINT centre to explore rater reliability across countries and support shared understandings of the protocol. These workshops proved very useful in strengthening confidence in correct scoring of lesson segments in accordance with PLATO. The scores for IC and CD in Iceland received special care and review as they were of central importance for the dissertation work.

A distinction worth emphasising in connection to observation analysis is between teachers and teaching. It is difficult to argue that "quality" is something that a teacher "possesses" or not. Rather, teaching quality is something teachers show with their actions during lessons. Therefore, the object of observation scores is teaching, i.e., the actions of the teachers, and not the teachers themselves. Furthermore, a teacher can show a great deal of quality in one lesson and a lesser degree in the next, as well as differently across segments within a lesson. It is not reasonable to expect teaching to be constantly at a high-level, since actions are contingent upon the purpose of each lesson and within the activity expected of students in each lesson segment. For this reason, the distribution of segment scores, showing presence or non-presence of high-level scores, give more information than plain computed averages.

For reporting results of these analyses, the observation evidence was used to categorise the mathematics classrooms into three groups: Group A, with evidence at the 4-level in either IC or CD (strong, consistent evidence); Group B, with the highest segment score in IC and CD being at the 3-level (evidence with some weaknesses); and Group C, where IC and CD scored consistently at the 1-level or 2-level (consistently limited or no evidence). Within groups, classrooms are ordered by the sum of their mean scores in IC and CD. The within-groups ordering is not particularly meaningful for group C, but it does distinguish between the evidence shown in classrooms in groups A and B. Furthermore, I argue that qualitative descriptions that clarify the arguments for the scores give greater credibility to the quantified segment scores. Qualitative descriptions are presented from each group that aimed to present a clear argument for how the teacher's actions resulted in their segment scores and therefore the group that they were placed in.

3.5.2 Cognitive activation through mathematical task analysis (TAG)

The potential for cognitive activation is often created through selection of tasks. Commonly, tasks are the basis on which the teacher creates opportunities for learning. Tasks with greater cognitive challenge may to a greater extent invite implementations that activate student thinking to build a richer understanding.

The Task Analysis Guide (TAG), proposed by Stein, Smith, Henningsen and Silver (2009; Smith & Stein, 2011; Stein & Smith, 1998), categorises tasks according to their level of cognitive demand. Four levels of cognitive demand are defined in TAG: memorisation and procedures without connections (low cognitive demand); and procedures with connections and doing mathematics (high cognitive demand). These four levels provide a framework for analysing tasks as they appear in curricular or instructional materials, the first phase of the Mathematical Tasks Framework (Stein et al., 2009). Tasks of low cognitive demand are algorithmic in nature and focused on producing correct answers without any explanations or reasoning of solutions. Tasks of high cognitive demand often have no pre-determined solution path and may involve multiple representations, requiring students to engage with conceptual ideas. The full rubric for the four levels in TAG is shown in Appendix B. Because of the emphasis on individual student work in textbooks in mathematics lessons in Iceland, the analysis of tasks selected for students to work on in the lessons provided further insight into the potential for cognitive activation, or specifically, the richness of mathematical practices.

The task analysis was carried out by me, using TAG in the same way I have applied it in previous studies (Sigurjónsson, 2014; Sigurjónsson & Kristinsdóttir, 2018). As I was the only rater of the tasks, there is no inter-rater agreement to speak of. TAG is free to use for teachers and researchers alike without a certification process. The results complement the PLATO observation scores as part of results addressing research question I.

3.5.3 Student perceptions connected to observation scores

To identify student perceptions of cognitive activation, specific questions from the Tripod survey were chosen. Since the intention was to explore the connection to observation scores in the specific dimension of cognitive activation, pre-defined categories in the seven C's were not suitable. Two subscales were constructed from six Tripod items that fit the specific aspects of teaching that are measured in PLATO. One subscale labelled reasoning formed a measure of student perceptions of to what extent the teacher engaged students' reasoning and explanations of their answers. The other subscale, labelled discourse, formed a measure of student perceptions of discourse in the classroom, i.e., the teacher respectfully inviting students to share their thoughts or ideas. The construction of the subscales may reveal possible nuances between similar constructs in PLATO and the Tripod survey. The reasoning scale was constructed from three survey items that best fit the IC element in PLATO. The discourse scale was similarly constructed from three survey items that best fit the CD element in PLATO. The reasoning and discourse subscales included specific items from Tripod's challenge dimension and confer dimension, respectively. **Table 8** shows the survey items used. A reliability analysis of the subscales yielded a Cronbach's alpha value of 0.67 for the reasoning scale and 0.65 for the discourse scale, which indicates acceptable internal consistency with three items in the scales. Intraclass correlations of the scales and intercorrelations are found in Appendix D. The mean score for all items in the survey was 3.63. To connect to the observation scores, the student responses were analysed with respect to the groups of teachers according to observation evidence, explained in chapter 3.5.1.

Table 8. Items used from the Tripod survey and Cronbach's alpha values for the two subscales.

Item	Item text	Subscale	α
REAS1	My teacher asks questions to be sure we are following along when s/he is teaching.		
REAS2	My teacher asks students to explain more about the answers they give.	Reasoning	0.67
REAS3	My teacher wants me to explain my answers—why I think what I think.		
DISC1	My teacher wants us to share our thoughts.		
DISC2	My teacher gives us time to explain our ideas.	Discourse	0.65
DISC3	My teacher respects my ideas and suggestions.		

It is worth noting that some of the survey items refer to the teachers' actions toward the entire class ("we" and "us") while others refer only to the specific responding student ("me" and "my"). In the Icelandic translation of the survey, this feature was retained, preserving some potential to reflect differences between how the students perceive the teacher acting toward them individually. Standard deviations within classrooms can be used as a proxy for student agreement in their perceptions. On a 5-point scale this number has a maximum of ~2.00 (half responds "never", and half responds "always") and a minimum of 0 (all students give the same response).

During data entry, there were some cases of ambiguous responses in the survey, and one type was handled systematically. In the case of a student making an adjacent double-cross on a single item, the response was determined with a randomised number generator service to randomly distribute the adjacent double-crosses in either direction ("RANDOM.ORG – True Random Number Service", 2022). The other, less common types of cases were a triple-cross on an item, and a non-adjacent double-cross. These responses were marked as invalid. Since the frequency of these ambiguous answers was distributed between many items, i.e., not systematically more frequent for specific items, further analysis of these ambiguous answers was not deemed necessary.

3.5.4 Qualitative analysis of selected Nordic lessons

To explore research question III required analysis of selected lessons from the Nordic video database with the strongest evidence of cognitive activation. The primary focus was on enriching empirical understandings of instructional formats and how cognitive activation is manifested in teacher-student interactions during lessons – given the fact that the selected lessons are the ones with the highest segment scores in cognitive activation across this Nordic dataset. Since there was no intention to draw conclusions about countries, it is not accurate to speak of a comparative study or comparative research. Rather, it is a qualitative inquiry aimed at enriching empirical understandings of what characterises lessons considered cognitively activating in a Nordic context. This inquiry involved both content analysis of the instructional formats in the lessons and reflexive thematic analysis of the teacher-student interactions.

The instructional format of the lessons was analysed using content analysis with a directed approach (Braun & Clarke, 2013; Hsieh & Shannon, 2005). The video-recordings and transcripts were used to map the design of the lessons on a minute-by-minute basis using the pre-directed categories: individual work, group-work, whole-class discussion, and direct instruction. Lesson time used for administrative tasks (e.g., taking attendance) or other non-math related matters (e.g., discussing Eurovision) was categorised as downtime. The pre-defined categories were the same as used for data logs in the QUINT database. However, in the data logs the most common format was coded at the segment level (15 minutes), i.e., not mapped on a minute-by-minute basis. The visual presentation of these results was inspired by Mok & Lopez-Real (2006). The content analysis also involved identifying tasks used in the lessons to show the mathematical topic of each lesson and what students were assigned to do.

I used reflexive thematic analysis to explore patterns of meaning across the teacher-student interactions (Braun & Clarke, 2022). While Braun & Clarke (2006, 2022) provide a generic description of the phases of reflexive thematic analysis, how this process looks like in practice for each study is non-standard. For my research process, I adopted a critical realist approach to the data. This allowed for including elements of inductive orientation to the data although the selection process was deductive in nature and, in a way, had directed my thinking as a researcher toward a deductive orientation. This flexible approach aligned well with the research aim, making the analysis empirically focused, although informed by existing theory. This move from purely deductive coding using PLATO to a highly flexible analytic process certainly entailed a vastly different approach to observing the lessons. To exercise the freedom of generating my own codes instead of using a pre-defined codebook was a challenge and required critically interrogating my interpretation of what the teachers said and did.

Before going on to describe how I implemented the phases of the reflexive thematic analysis, it is worth noting how I came to the decision to employ the method. Initially, I intended to apply some pre-existing framework to deductively analyse the tasks found in these lessons. However, upon examining the data, I soon grew distant from the idea. I felt it would be more informative and useful for addressing the research questions to move the focus from a task analysis to a focus on the teachers' implementation of the tasks in interaction with students. In the results, I show the tasks that appear in the lessons within an overview of the instructional formats that resulted from the content analysis. Ultimately, the choice of reflexive thematic analysis of interactions allowed me to develop themes that built on codes that were partly inductively generated from the data and partly deductive as based on my own knowledge and understanding of teaching practices. The conclusions also involved abductive reasoning, which combines theory-driven codes and inductive reasoning (Patton, 2015). This flexibility in coding was important in developing themes that, through abductive reasoning, focused on enriching empirical understandings of cognitive activation as a theoretical concept.

An integral part of reflexive thematic analysis is a continual critical examination of the research process. For my research process, essential to this examination was regular writing in a reflection journal throughout the entire process, storing and documenting thoughts for meaning-making of the data and decisions made throughout. Braun & Clarke describe reflexive journaling as "one of the most important practices you'll undertake on your research journey" (2022, p. 19). I would argue that reflexive journaling was the most important practice throughout all six phases of my research process.

In the first phase, the aim was to familiarise myself with all the lessons. I watched all lesson videos from start to finish with the translated transcripts on a second screen (the lesson videos from Sweden and Norway also had English subtitles). In this phase, I wrote notes in a separate document on the content and structure of the lessons and marked each interaction event in the transcript. The distinction between interaction events was based on two premises on teacher-student interactions: A teacher can interact with a single student, a small group of students, or the entire class, and interactions can be on different topics. Thus, the distinction was made if one of two criteria was met: 1) the teacher turned the interaction toward another person, or another group of persons, without including the person(s) previously addressed, or 2) the teacher changed the topic of the interaction. In addition, to provide necessary context in the transcript, each interaction was specifically marked to specify whether the teacher's words were directed at the entire class or a subset of the class, such as a small group or an individual student.

The next phase of analysis was the coding phase. I imported the interaction-marked transcripts into the Atlas.ti computer software where I coded each interaction. As mentioned, the coding was open with no pre-defined codes, though certainly informed and shaped by my personal experience and knowledge of teaching and theories in education. I began to form clusters of codes during the coding phase, as I observed some similar types of interactions across the lessons and interpreted common motives for these interactions. To mark these clusters of codes, I used functionality within the software: colour-coding and code groups.

I envisioned that the coding process would be based on my own perception and interpretation of how the teachers interacted with the students: what was said and how, who did the thinking, which questions were asked and how, how students were assisted, how the mathematical content was framed, and what students were asked to do. I also speculated that I could consider other aspects that were not as connected to my knowledge of teaching quality, e.g., to what extent these lessons were consistent with a "Nordic model" of teaching, and what I could notice in the students' expression, such as joy or satisfaction with the lesson. These speculations did not end up as codes in the data, as I perceived that the data at hand to not invite such interpretations. However, the results may be readily discussed in connection to prior research on Nordic mathematics teaching (as reviewed in chapter 2.1). I also felt more strongly that the analysis should be without the "PLATO-glasses" in interpretation of the data. I preferred to let PLATO suffice as sampling criteria and nothing more. However, it was unavoidable for me to be conscious of my knowledge of the observation scores in analysing the lessons. As informed by my critical realist approach, the initial codes, although data-driven, were still inspired by my knowledge of prior research on teaching and learning. Not all of the initial codes ended up as part of themes in the results, which is normal for reflexive thematic analysis (Braun & Clarke, 2022). Examples of a coded data extract is in Appendix F.

In the theme generation phase, I used the clusters of codes to generate themes that could meaningfully contribute to addressing the research question of what characterised teacher-student interactions in the lessons. This involved writing out how each candidate theme may address the research aim of enriching empirical understandings. Around the time that I moved toward the theme development phase, I had (finally, after pandemic-related delays) arrived in Karlstad, Sweden, for my stay abroad as required by my university's doctoral program. My own personal interactions with other researchers there were very useful to develop my interpretation of the data. I had many fruitful discussions about my initial candidate themes with Jorryt van Bommel, expert on my doctoral committee located at Karlstad University, as well as her doctoral student, Jimmy Karlsson (whose research topic is also cognitive activation). As I moved into the next phase, conversations and feedback at research seminars and a conference presentation connected to my stay there helped to refine, name, and define the themes.

After arriving back home in Iceland, I entered a second round of coding the lessons, in a reverse order from the first round. In this second round of coding, I already had some candidate themes that had been developed after presentation and discussion at the various seminars and meetings in Sweden. It was a challenge to not become too attached to those initial candidate themes, but even though I considered them relatively well-defined at that point, they were subject to change. The second round tested the applicability of the developed candidate themes and further refined them. The writing phase started in Karlstad, interwoven with other phases from theme development, and concluded with revisions in my usual workspaces in Iceland. Appendix G illustrates the development from initial themes to the final themes presented in the dissertation's results.

3.6 Ethical considerations

Ethical issues are inherent in all research and addressing them is especially relevant with video data of classrooms. Issues of importance include informed consent of all participants, how data are stored and handled when it comes to privacy and anonymisation, and how results are presented.

An informed consent form was introduced to and received from all participant teachers, each student, and their guardians in the recorded classrooms by the following process:

1. Selected schools were contacted by phone and e-mail, first the principal and then the participating teachers, with a brief description of the project. If teachers agreed to participate, they gave informed consent with their signature on the form. If one or more of the teachers declined participation, another teacher or a new school was contacted.

2. The project was introduced to students' legal guardians. Principals made the first contact by sending an information letter. A member from the research group was available to attend a meeting with guardians to present the project upon request. A form for informed consent was signed by each participating student's legal guardian. In the case of guardians choosing not to accept participation, a special arrangement was made regarding the student to attend the lesson in a different setting within the school while the data collection took place.

3. The project was introduced to the students by a teacher. The students filled out the same consent form as their legal guardians.

The consent form prompts agreement or disagreement to different parts of the study, such as being seen in the videos, use of videos in professional development and teacher training, and if the videos can be shared with other researchers within QUINT (see Appendix E). Their preferences to each of these different parts are carefully documented and fully respected. Participants are informed that they can withdraw from the study at any time.

The identity of participants, both teachers and students, is kept anonymous in all presentation of results. In the presentation of the Iceland data, teachers are strictly referred to using indexed letters, i.e., T_1 through T_{10}. This numbering is based on the group ordering by observation evidence explained in chapter 3.5.1 and illustrated in the findings (chapter 4.1.1). In the qualitative analysis of the eight Nordic lessons, the two teachers in each country are given pseudonyms starting with the same letter as the name of the country in which their lesson was recorded, as further outlined in chapter 3.3 on participants.

The Iceland classroom video data are securely stored at The Educational Research Institute (Menntavísindastofnun). They are also stored in a secure video library at the University of Oslo. Access to the video library is only granted through a special permission (for details, see Klette, 2022). All procedures about confidentiality, permission, storing and sharing of data have already been acknowledged by applicable authorities in Denmark, Norway, and Sweden. In Iceland, the study is conducted in accordance with the Data Protection Act no. 90/2018 and has been reviewed successfully by the Science Ethics Committee of the University of Iceland.

3.7 Limitations and strengths

The research project set out with the potential to create new knowledge about mathematics teaching practices and teaching quality in Iceland and other Nordic countries through video-recordings of lessons and a student survey. Both approaches have strengths and limitations.

A strength of this research approach is the in-depth view it provides into Nordic mathematics classrooms. This is realised both through a deductive systematic approach of lessons from Iceland (i.e., PLATO coding) and a more qualitative, inductive analysis of specific Nordic lessons that may be considered outstanding. This allows for both rich descriptions of what "goes on" in Nordic classrooms with high cognitive activation, as well as to what extent mathematics teaching in Iceland, as observed from these ten teachers, can be considered cognitively activating.

A first limitation has to do with the size of the sample. Video data are notoriously time-consuming to analyse (see, e.g., Blikstad-Balas, 2017). Therefore, video studies do not typically sample more than a few dozen schools, a small fraction of a given school system. This limits inductive claims and generalisability. In Iceland, around 4500 children attend primary school at each grade level (Statistics Iceland, 2020). This study reached 217 students in grade 8 mathematics which is approximately 5% of all students in grade 8 at that time. From each classroom, data was collected from three to four lessons in mathematics, equivalent to approximately one school week of math. Since the school year in Iceland is 180 school days, one school week is approximately 3% of the school year. To make inductive claims about all mathematics teaching in Iceland would require the sample to be highly representative. Even though the schools were sampled to capture breadth in school characteristics, such as school size, location, and different proportions of students with immigrant backgrounds, such claims are difficult to fully justify. Forming the research questions required me to consider these limitations. For instance, for research question III it was chosen not to explore a direct comparison between countries, but rather an inquiry into carefully and specifically selected "high-quality" lessons in the LISA Nordic dataset. Although a strength is in the detailed results about cognitive activation teaching practices in a Nordic context, there are severe limits to conclusions that can be drawn about teaching in each country as a whole.

Another limitation is in the observational instrument, PLATO. Use of this observation instrument was pre-decided by the QUINT research initiative. Participation in the analysis of the dataset, attending a course to be trained and attending various workshops for rater calibration efforts was a great learning opportunity for me and developed my thinking about not just the observation system but to different perspectives of teaching quality as well. However, it may be mentioned as a limitation to the study that PLATO is only one perspective of teaching quality. If I had the resources to research and choose my own instrument, it is possible that the results would be differently phrased or skewed to other emphases. Furthermore, the transferability of results from PLATO coding is limited by the general reductionist nature of observation systems (Blikstad-Balas, 2017). The way the IC and CD elements in PLATO are conceptualised is one way of measuring cognitive activation through lesson video data, and it is difficult to directly compare results to other ways of measuring cognitive activation. Every observation instrument has limitations in both how it measures and what it detects as aspects of teaching quality. Placing trust in raters' scores was another inevitable limitation, as rater error is a known issue in systematic observations (White, 2018). However, as previously described, rater error in the analysis of QUINT data was addressed with double coding and calibration efforts through video workshops. Although I had to place trust in PLATO codes from other countries for selecting lessons, it did not pose limitations for results based on what I observed in the lessons.

For the qualitative analysis there are both limitations and strengths. A strength of the method is in the detailed and in-depth analysis of carefully selected lessons. It fits well with the aim of developing rich understandings of lessons with high cognitive activation. However, the interpretive approach is non-standard, limiting transferability. When it comes to comparisons between different lessons in cognitive activation, comparisons can be drawn from the descriptions of different groups of teachers according to observation evidence according to PLATO.

Lastly, a potential limitation is the small amount of information about the student groups or individual students. This is a result of the research design to focus on the teachers, and it can be partly explained by issues of research ethics it might raise. However, certain teacher actions may be based on responding to learner needs that the researchers are not specifically aware of. A teacher may often focus on the individual students that they are teaching and not consider the whole group of students in the same way as is done in the study. This has possible implications for research question II. The study explores the connection between one student survey and one classroom observation system. To what extent the results generalise to other instruments is not fully known.

3.8 Summary

This chapter has outlined the methods used to reach the aims of the doctoral research project. QUINT's design of a video study with student surveys and additional contextual data, e.g., in data logs, was explained and how it allowed for the research design of this research project with a focus on cognitive activation. The process of video-recording lessons from ten schools with "naturally occurring" teaching was explained. A description of how mathematical tasks were identified from both video and contextual data followed and the collection of student perception data through the Tripod survey. The procedure used to select outstanding lessons in cognitive activation from the Nordic dataset based on PLATO scores was explained.

Analyses of these data were then described. The use of PLATO to systematically generate results on cognitive activation through the four levels of IC and CD were explained, as well as the analysis of mathematical tasks into four categories of cognitive demand. The specific items and subscales pertaining to reasoning and discourse from the student survey were then clarified. The qualitative inquiry of the selected Nordic lessons, through reflexive thematic analysis of interactions and content analysis of instructional formats, was explicated in detail. The chapter then concluded with a discussion of various ethical considerations inherent to collection and analyses of these data, as well as a discussion of the limitations and strengths of these methods.

4 Findings

An extended summary of the main findings of the research project is presented in this chapter, organised in sub-chapters by the three research questions. The overarching aim of the research project was to develop a deeper understanding of cognitive activation in mathematics in Iceland and in a Nordic context. In chapter 4.1, findings are presented that describe cognitive activation in mathematics teaching in ten classrooms in Iceland and the potential for cognitive activation through mathematical tasks (Aims 1 & 2). In chapter 4.2, the connection between observation scores and student perceptions of cognitive activation in these ten classrooms is described (Aims 3 & 4). In chapter 4.3, the results of the qualitative inquiry of interactions and instructional formats in eight outstanding lessons in cognitive activation from the Nordic dataset are presented (aims 5 & 6).

4.1 Cognitive activation in mathematics teaching in Iceland

The analysis of the video-recorded mathematics lessons in Iceland showed varying levels of evidence for cognitive activation between classrooms. The first half of this section is devoted to describing different groups of classrooms according to observation evidence, providing examples of lesson episodes and reasoning behind the segment scores that exemplified each group. The task analysis showed mostly procedural tasks without connections to mathematical concepts or meaning, but also procedural tasks with connections and "doing mathematics" tasks. The second half of the section provides further delineation of the findings from the task analysis and puts them in a perspective with the teaching observation findings.

4.1.1 Three groups of mathematics classrooms by observation evidence

As illustrated in **Figure 6**, observation scores were mostly on the low end (1-level or 2-level) for both IC and CD. **Table 9** shows the grouping of teachers into three groups according to observation evidence: Group A, with evidence at the 4-level in either IC or CD; Group B, with the highest segment score at the 3-level in IC and CD; and Group C, where IC and CD scored consistently at the 1-level or 2-level. Within groups, teachers are ordered by the sum of their mean scores in IC and CD.

Table 9. Ordering of teachers by ordinal groups from maximum scores in IC and CD, and within groups by sum of mean scores of IC and CD.

Teacher	max(IC)	max(CD)	Group	mean(IC)	mean(CD)	mean(IC+CD)
T_1	4	4	A	3.50	3.38	6.88
T_2	4	4		2.18	2.27	4.45
T_3	3	3	B	2.60	2.40	5.00
T_4	3	2		2.56	1.67	4.22
T_5	3	2		1.55	1.27	2.82
T_6	2	2	C	2.00	2.00	4.00
T_7	2	2		1.80	1.60	3.40
T_8	2	2		2.00	1.38	3.38
T_9	2	2		1.75	1.38	3.13
T_{10}	2	1		2.00	1.00	3.00

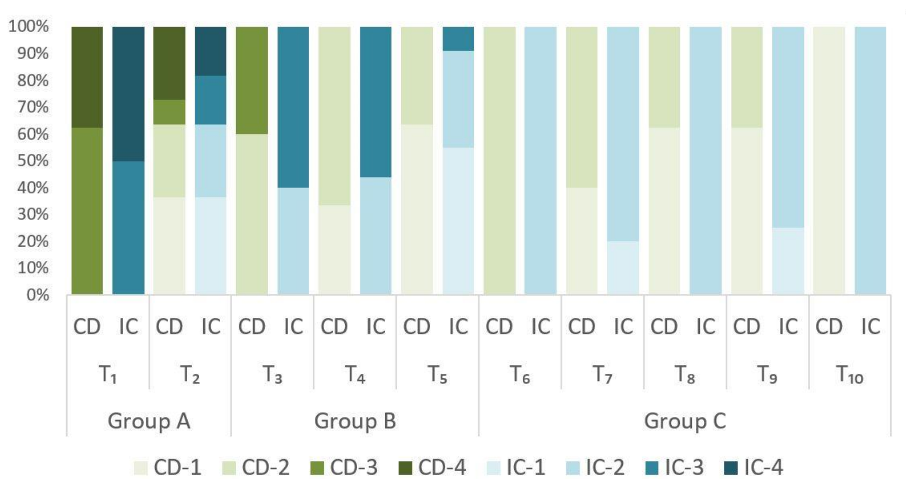


Figure 6. Distribution of PLATO scores for IC and CD in Iceland across classrooms with observation evidence groupings.

The segment scores of the observed lessons show substantial variation between teachers in the level of cognitive activation in their lessons. However, this quantitative variation does not describe in detail how this is manifested in instructional practice. To paint a rich picture of this difference in practice, some illustrative descriptions from each of the three groups are presented.

The two teachers in group A showed consistent evidence of high cognitive activation, but in different ways. T_1 (Íris in chapter 4.3) showed consistent evidence at the high end (3-level or 4-level) in IC and CD in all observed segments. Her lessons began with brief review with the whole class. Within minutes, students were put to work on a specific task. They were assigned to work in pairs and instructed to explain their thinking to each other. This resulted in high-level scores in both IC and CD from the first segment of each lesson. In one lesson, students had the task of using dice to make a probability experiment. They were to document the results and compare them to the theoretical probability. When students had questions, the teacher would sometimes pick up on those questions on the whiteboard. For example, when asked whether the order of numbers matters in multiplication, the teacher showed, with student input, an example of the commutativity of multiplication. Even when modelling a tree diagram for students on the whiteboard, the teacher explanations built on questions of why and how with input from students. This was in harmony with the discursive nature of the lesson. The following excerpt is in her assistance to a student pair as they were working on the dice-throwing probability task:

> T_1: What is the probability of getting a five if you throw a dice?
> S_1: One over six.
> T_1: One over six. How often should I get a five if I would throw the dice 30 times?
> S_1: 30, isn't that... five?
> T_1: Five times. Why?
> S_1: No six times, because five times six is 30.
> T_1: Do you agree with that? Mhm. Ok.
> S_2: I don't get it at all!
> T_1: S_1, can you explain to S_2?

This interaction contributed to the argument to advance the IC score from the 3-level to the 4-level. By weaving together a rigorously successful implementation of demanding tasks, uptake of student ideas and ample opportunities for mathematics-related classroom discourse throughout her lessons, the teaching observed from T_1 was consistently scored at high-levels in both IC and CD.

T_2 (Ída in chapter 4.3) had more variation in segment scores. Her lessons started by instructing each student to either silently watch an instructional video (using headphones on their tablets) or listen to a short lecture from the teacher at the whiteboard. These opening lesson segments resulted in scores at the 1-level in IC and CD. However, as the lessons went on, the segment scores in these elements trended upward. Similar to T_1, T_2 commonly assigned students to work in pairs or groups of three. In one lesson, the task was to come up with mathematical expressions that their peers were then to simplify by correctly using the order of operations. Toward the end of the lesson, the teacher asked students to explain their steps toward a solution and wrote their responses on the whiteboard. This type of instruction created rich opportunities for student talk, uptake of student thinking and ideas, and it invited students to reason and justify their solutions. At the end of the other lessons, T_2 invited students to a whole-class activity in the math game called Twenty-four. Four numbers were written on the whiteboard and the students were to use mathematical operations to create an expression with the value 24. This activity invited opportunities for mathematical discussion among students and intellectual challenge through explaining their thinking process, resulting in high-level scores. Although there were examples where T_2 diminished intellectual challenge through overly directional comments (e.g., "Perhaps multiply by two") or telling students the solution to tasks (e.g., "Yes, then y is equal to 3x here"), in other segments she maintained and even advanced the challenge. The rich opportunities for classroom discourse and student reasoning typically found at the end of the lessons resulted in scores at the 4-level, placing T_2 in the group A with some strong evidence of cognitive activation.

In group B, T_3 had segments scoring at the 3-level in both elements. In an introductory lesson about equations, students were first shown picturesque equations on a projector with everyday objects (e.g., fruits, Pokémon) to signify variables before moving on to equations using conventional alphabet-letters such as x and y. Students were instructed to spend a couple of minutes to discuss solution strategies among peers sitting by their tables and then to communicate their results while the teacher elaborated by the projector. This involved some high-level uptake from the teacher, but typically more teacher-directed discourse, resulting in CD at the 3-level. T_3 would usually maintain the challenge of the tasks, keeping IC at the 3-level. However, in the last segment of the lesson the teacher presented students with a game. Equations appeared on the projector with four solution options to which each student responded in their own tablet. The students had limited time to respond to each equation. Once the time was up, the distribution of responses to each option was shown on a graph. Although the activity seemed somewhat promising for cognitive activation, it ultimately did not facilitate justification or mathematical discussion between the students and seemed more geared toward individual competition in solving equations quickly. This segment scored at the 2-level in IC and 1-level in CD.

Alongside T_3 in group B were T_4 and T_5 which scored at the 3-level in IC but never above the 2-level in CD. T_4 scored at the 3-level in IC in roughly half the observed lesson segments. Teacher T_5 only had one segment at the 3-level in IC. These teachers kept their students active individually in textbook activities but did not specifically instruct students to collaborate or engage in mathematical discourse, resulting in CD being consistently at the 1-level or 2-level. T_4 would at times show uptake of student ideas at a high-level, but, with extended opportunities for student talk lacking, CD remained at the 2-level. In their assistance to students, both teachers would sometimes reduce the challenge of the task at hand. This was especially common with T_5. However, T_4 was more persistent in eliciting student thinking in conversation while assisting individual students, resulting in IC at the 3-level in some segments. The following excerpt is from T_4's assistance to a single student working on a task applying the distributive property[f]:

> S: How do I calculate this here?
> T_4: How do you calculate this – for example, what did you do here?
> S: I just did times three.
> T_4: You just do the same here, take the numbers and add.

This interaction neither reduced nor advanced the challenge of the task. The teacher merely pointed the student toward a previous solution that he could build on and extend to a calculation with different numbers and letter variables. The full task at hand in this excerpt is shown in **Table 11**.

Group C consisted of teachers with limited evidence of cognitive activation, i.e., no segment scores above the 2-level in either IC or CD. Their lessons had some shared characteristics. Students typically worked individually and often "at their own pace" in textbooks. This meant that not all students were assigned the same specific tasks to work on at the same time. Further, they were not specifically instructed to collaborate with their peers. Usually, the teacher would use the lesson time mostly to circulate the room and assist students individually. In their assistance, the teachers would often reduce the challenge of the tasks by asking closed questions or telling students the solution, commonly resulting in IC scores at the 2-level. For this reason, teachers T_6, T_8 and T_{10} scored at the 2-level in every lesson segment. The following excerpt is from T_{10}'s assistance to a single student working on a geometrical task. The task is to construct a circle with the diameter AB and a circular arc AC of 135 degrees, and then calculate angle values of the triangle ABC:

> S: I don't quite understand.
> T_{10}: Draw AB on the picture.
> S: Which one should it be, AB or a b.
> T_{10}: That doesn't matter. Then you draw a segment AC here between a and c, here between. Were you gonna make the line the other way?
> S: But this is c.
> T_{10}: Yes, this is the line c, but this is the point C. Well, you were allowed to put... so you are done drawing the diameter, mark here again on the line and stop this nonsense. Then put a segment, and then a new segment like this, okay, and now you have it, a, c, b, and calculate the corner ABC, this angle here. This angle is 135 degrees, mark it in here. No, not that way, this way, 135. How many degrees are left in the triangle?

---

[f] Which particular expression the student asks about is unclear, but the group of expressions in the textbook exercise is shown in Table 11.

S: 45.
T_{10}: Yes, 45, and because this side is the same length as this one then this is 22 and a half degree, no not on that side, inside the triangle.

This interaction was one of many that were considered to reduce the challenge of the task, resulting in the score in IC being lowered from a 3 to a 2.

Some teachers in group C would occasionally pick up on student ideas on the whiteboard or facilitate brief discussions in their assistance to students. However, this discourse was never brought to the whole class and typically had to do with applying procedures where the goal was to find the solution, as opposed to develop an understanding of the associated mathematical concepts. This instructional decision meant that scores remained at the 2-level. The exception to this was T_{10} which did not show any evidence of uptake of student ideas. This meant CD was scored at the 1-level in all observed lesson segments from T_{10}.

4.1.2 Tasks: Distribution, volumes and potentials

The result of the task analysis by classrooms and groups according to the observation evidence is shown in **Table 10**.

Table 10. Number of tasks in each TAG category across classrooms.

Classroom	Group	TAG-1	TAG-2	TAG-3	TAG-4	Total
T_1	A	0	5	4	1	10
T_2		0	4	1	1	6
T_3	B	0	1	9	0	10
T_4		0	9	1	0	10
T_5		0	14	2	0	16
T_6	C	0	5	4	0	9
T_7		0	5	2	0	7
T_8		0	12	9	1	22
T_9		0	8	3	0	11
T_{10}		0	30	12	1	43
Total		0	93 (64%)	47 (33%)	4 (3%)	144

Out of 144 tasks analysed across the ten classrooms, no tasks were identified in the Memorisation category. This indicates that neither textbooks nor teaching directly aim to have students memorise mathematical facts. However, 64% of tasks were in the TAG-2 category: procedures without connections. TAG-3, procedures with connections, counted 33% of all tasks. Only 4 tasks out of 144, or 3%, were identified in TAG-4, doing mathematics. These results suggest that most tasks are aimed at procedural fluency. Though some tasks aim at conceptual understanding, it is to a large extent up to the teacher to support that development.

Four different teachers had a task identified in TAG-4, two from group A and two from group C. In group A, T_1's task came from a textbook task of making a probability experiment, and for T_2 it was in a game of Twenty-four (both discussed in the previous sub-chapter). In group C, both tasks came from textbooks. T_{10}'s task had to do with exploring two different methods of constructing an increasing number of triangles from sticks, creating different formulas for each method, and exploring boundary problems if only 100 pins were available for use. T_8's task is illustrated in **Table 11** along with examples of tasks in the data from each category.

Table 11. Proportion of tasks in each category and task examples.

TAG category	Proportion	Task example
Memorisation	0%	N/A
Procedures without connections	64%	Use the correct rules of operations to multiply together the numbers or letters with the expressions in brackets. Write the solution as simplified as possible. a) 3(a+1) d) m(10−3) b) (4+b) · 2 e) (2+4x) · y c) (x−6) · 9 f) (3s+5t) · 7
Procedures with connections	33%	This list shows the number of goals that the handball team Heroes scored in 20 games last season: 24, 15, 24, 21, 19, 12, 14, 21, 27, 24, 24, 18, 26, 20, 23, 21, 21, 27, 18, 24 a) Categorise the data into groups with a width of 5 and find the frequency in each group. Write your results in a table. b) Create a histogram from the data in a) c) Categorise the data again, this time into groups with a width of 3. Find the frequency in each group and create a histogram. d) Which histogram do you think gives a better overview of the data?
Doing mathematics	3%	A class has 20 students. Nine students have a sister and 10 have a brother. Five students have no siblings. How many students have a brother and a sister?

A closer look at the task analysis within different groups of teachers according to observation evidence yields some findings with regards to not only the distribution but also the volume of tasks. In group A, there was a relative balance between purely procedural tasks and tasks aiming at deeper mathematical connections. T_4 and T_5 in group B tended toward more procedural tasks, while T_3 had more tasks with connections to concepts. In group C, all teachers had a majority of their tasks in TAG-2: procedures without connections. Most teachers assigned around 10 tasks across the lessons, but some assigned more. T_5 had very fast-paced lessons and amounted to 16 tasks. T_8 and T_{10} amounted to 22 and 43 tasks, respectively. The reason for this high number of tasks was that these two teachers had students working at their own pace in different parts of the textbook. For T_{10}, tasks were identified from three different chapters (statistics, geometry, and algebra) across two textbooks. Since students in these lessons were working on different parts of the textbook at the same time, the number of tasks identified went up significantly.

4.2 The connection between observed cognitive activation and student perceptions

Before describing the connection between student perceptions and observation scores in mathematics in Iceland, some descriptive statistics on the student responses are presented. **Table 12** shows the number of responses to each item along with the overall mean, median, standard deviation, and skewness. Every item was negatively skewed, with response to the items more likely on the favourable side than non-favourable. **Figure 7** illustrates this with histograms of the overall response to each item (where 1=never, 2=seldom, 3=sometimes, 4=often, and 5=always).

Table 12. Descriptive statistics of overall response to the specific Tripod items.

	REAS1	**REAS2**	**REAS3**	**DISC1**	**DISC2**	**DISC3**
N	210	212	210	206	206	208
Missing	7	5	7	11	11	9
Mean	3.67	3.23	3.48	3.03	3.39	3.51
Median	4.00	3.00	4.00	3.00	3.00	4.00
SD	0.985	0.968	1.02	0.952	0.971	1.03
Skewness	-0.443	-0.099	-0.341	-0.127	-0.244	-0.266

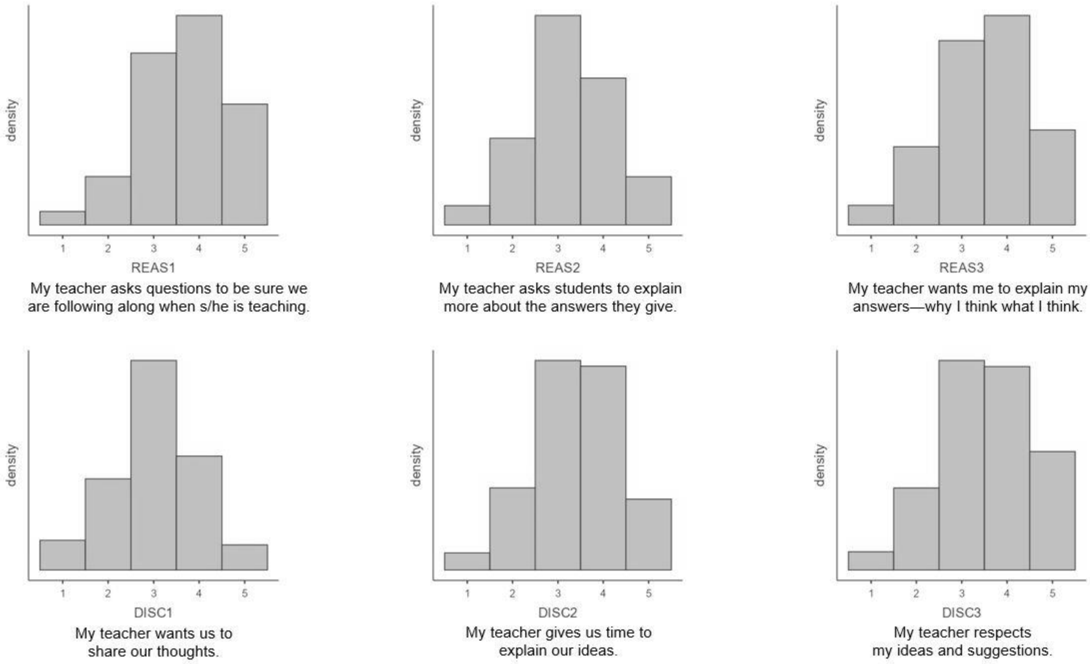


Figure 7. Distribution of overall student response to the specific Tripod items.

Compared to other items in the survey, these six items were toward the low end. All items except REAS1 had a mean lower than the overall mean for all items in the survey (M=3.63). The items REAS2 and DISC1 were particularly low, around half a standard deviation below the overall mean.

Table 13 shows the mean and standard deviation of the reasoning and discourse scales by classrooms in order by the observation evidence of cognitive activation (as further elaborated in chapter 6.1.1). Viewed in comparison with this grouping and ordering of classrooms by observation evidence, one can consider the connection between the two measures.

On the reasoning subscale, student ratings of T_2 stood out with the highest mean score and the lowest standard deviation, indicating high agreement among students that T_2 offered them opportunities to explain and reason for their answers and results. For T_1, with the strongest observation evidence, the mean score for reasoning was around or slightly below the mean with average student agreement. Student ratings in group B all hovered around the mean score in reasoning with average student agreement. In group C, T_6 stood out with the second highest mean student rating overall. T_{10} had the lowest mean student rating, indicating students did not sense opportunities to explain and reason for their results. T_{10} also had the second lowest student agreement, just above T_9. The other teachers in group C were close to the overall mean student rating.

Table 13. Means and standard deviations for the Reasoning and Discourse subscales along with PLATO observation scores for IC and CD.

		Reasoning		Discourse		Group	PLATO-IC			PLATO-CD		
Teacher	***n***	**Mean**	**SD**	**Mean**	**SD**		**Max**	**Mean**	**SD**	**Max**	**Mean**	**SD**
T_1	23	3.26	0.78	3.35	0.52	A	4	3.50	0.53	4	3.38	0.52
T_2	9	4.26	0.36	3.70	0.93		4	2.18	1.17	4	2.27	1.27
T_3	37	3.53	0.70	3.14	0.83		3	2.60	0.55	3	2.40	0.55
T_4	16	3.38	0.50	3.33	0.80	B	3	2.56	0.53	2	1.67	0.50
T_5	20	3.47	0.71	2.93	0.78		3	1.55	0.69	2	1.27	0.47
T_6	10	4.03	0.76	3.93	0.86		2	2.00	0	2	2.00	0
T_7	11	3.30	0.72	3.39	0.61		2	1.80	0.42	2	1.60	0.52
T_8	28	3.55	0.67	3.63	0.66	C	2	2.00	0	2	1.38	0.52
T_9	11	3.48	0.92	3.42	0.54		2	1.75	0.46	2	1.38	0.52
T_{10}	24	3.07	0.85	3.18	0.70		2	2.00	0	1	1.00	0
Total	189[a]	3.46	0.76	3.32	0.75			2.15	0.77		1.80	0.87

On the discourse subscale, T_2 received the second highest student rating, but in this case with the highest standard deviation, indicating low student agreement on whether they sensed time given for sharing their thoughts, ideas, and suggestions. T_1 received an average student rating, but here with the highest student agreement. In group B, student ratings were at or slightly below the mean rating with average student agreement for all teachers. T_6 had the highest student rating for discourse with relatively low student agreement. Teachers in group C all received a student rating close to or above the overall mean for discourse. T_{10} had the lowest rating among them, yet not as low as T_3 and T_5 in group B. **Figure 8** illustrates the mean student perception on the reasoning subscale by classrooms in order by the observed evidence of cognitive activation. **Figure 9** does the same for the discourse subscale. The dotted line shows the overall mean in each scale.

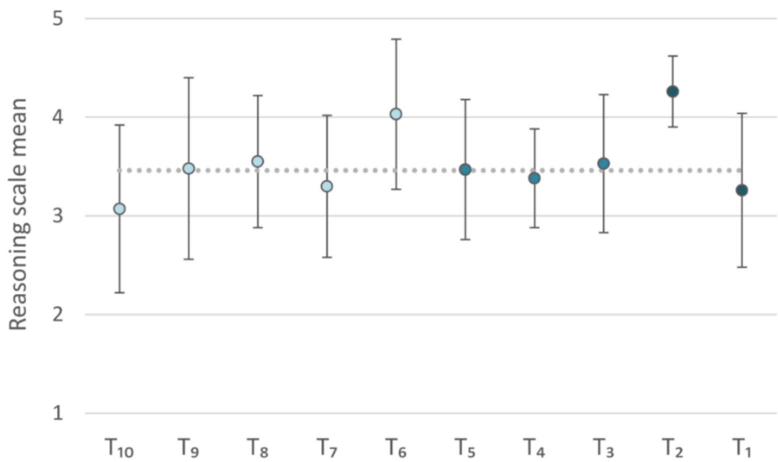


Figure 8. Reasoning scale means by classrooms in order by observed evidence of cognitive activation.

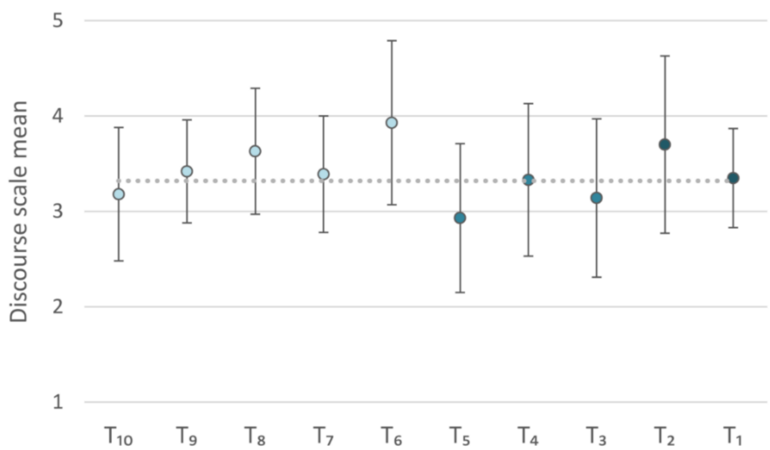


Figure 9. Discourse scale means by classrooms in order by observed evidence of cognitive activation.

As the figures illustrate, the findings indicate a weak connection between observation scores and student perceptions of cognitive activation. The variation of student perceptions was generally greater within classrooms rather than between them. As presented in Appendix D, rank correlations between the scales and the teacher order according to observation evidence are weak.

4.3 Understanding instructional formats and interactions in cognitively activating mathematics lessons in a Nordic context

This sub-chapter turns to the eight Nordic mathematics lessons specifically selected for further analysis based on their outstanding observation scores in IC and CD. For Iceland, this includes lessons from T_1 and T_2, which comprised group A (now using the pseudonyms Íris and Ída, respectively). I begin by giving an overview of the lessons, showing selected tasks, and explaining the variety of instructional formats found both within and between lessons as a result of the content analysis. Subsequently, I present and discuss the three themes on teacher-student interactions that were developed from the thematic analysis. The first theme describes different types of interactions seen in the lessons and how the teachers would often shift rapidly between them. The second theme describes the use of both formative feedback to monitor student progress and understandings, and explicit student roles to encourage content-related interactions between students. The third theme describes the connection-making that the teachers were observed to engage in with examples of how connections to both mathematical and non-mathematical experience seemed to be aimed at pushing students toward relational understanding.

4.3.1 Instructional formats varied within and between lessons

As illustrated in **Figure 10**, the tasks and activities teachers had selected for the lessons originated from a range of mathematical topics, such as algebra, probability, fractions, percentages, and division. None of the lessons were about geometry, in which argumentation and reasoning is often an implicit part. Typically, the lessons had few tasks that the class as a whole or in groups would engage in together. For instance, Sabrine's class only had two tasks, the candy-sharing task and the "flowerbed pattern" task (see **Figure 10**), each implemented in a think-pair-share type of format. Some lessons were more fast-paced but had a specific focus where the entire class commonly worked together. Examples of this are Doris' lesson on percentages and Nils' lesson with a focus on evaluating and expanding fractions. Some tasks can be described as mathematically rich, such as Sabrine's tasks and the game of Twenty-four toward the end of Ída's lesson. Other selected tasks in the lessons were procedural in nature, such as simplifying expressions (Ída) or carrying out a division procedure (Nadia). Consequently, there were other aspects than the selection of tasks that made these lessons cognitively activating.

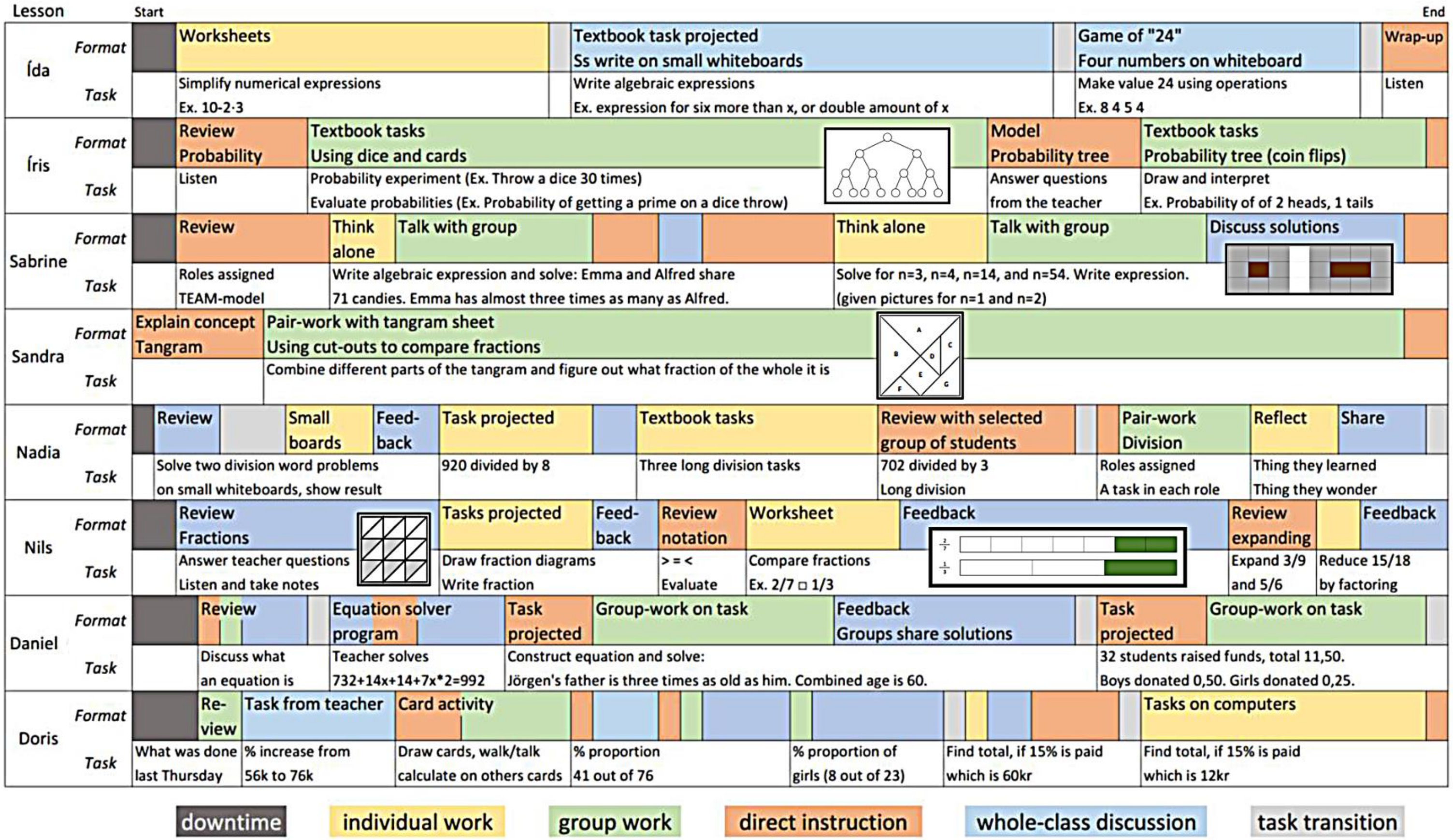


Figure 10. Timeline of instructional formats and tasks in the eight Nordic lessons.

As also seen in **Figure 10**, many lessons used visual models. Íris' probability lesson worked with probability trees, while Sabrine's algebra lesson used flowerbed patterns to explore algebraic patterns and formulas. Sandra's fraction lesson used tangram tiles to explore fractions as part of a whole, while Nils' fraction lesson used other geometrical models such as bar models to evaluate fractions.

The lessons were also differently placed in the learning trajectory. For instance, in Ída's lessons, students had their first introduction to algebraic expressions as an extension from numerical expressions and order of operations. Nadia's lesson aimed at revising division which students had learned about before in previous grade levels. Other lessons appeared to be toward the middle of a unit.

The instructional format of the lessons varied both within and between lessons. As also demonstrated in **Figure 10**, group-work (green colour) and whole-class discussions (blue colour) were formats that were observed in every lesson. Only two lessons (Ída and Nils) had no group-work, but instead had extended whole-class discussions. Similarly, only two lessons (Íris and Sandra) had no whole-class discussions, but instead spent most of the lesson time on group-work. Some lessons had individual work (yellow colour) but in every case it covered only a short amount of lesson time, and never extended for longer than 15-20 minutes before shifting to a different format. For example, in Sabrine's lesson, students were to think about the candy-sharing task alone for a minute before talking with their group as part of the think-pair-share format. All lessons had some direct instruction (red colour) but always over a short amount of time.

Segments of direct instruction often occurred in the beginning with review or toward the end for wrapping up. Direct instruction was also scattered in the middle between other formats, commonly including explanation of the next lesson activity, a brief review of concepts or explanation of content that the students were engaging with. For instance, in Nadia's lesson she had a few students come to the blackboard to review division with her as the other students continued their textbook exercises. As such, the lessons were not devoid of what may be called "traditional mathematics teaching" – parts of some lessons had procedural exercises and rules or examples on a blackboard to follow, or sections in which the students were mostly to sit passively and listen (i.e., not actively engaging with content). However, these sections did not extend a large portion of the lessons. The extended sections tended to be either group-work or whole-class discussions, sometimes in a game-like format (e.g., game of Twenty-four in Ída's lesson) or with students assigned explicit roles (e.g., Sabrine's and Nadia's lessons). Further, none of the lessons followed a particular pattern – all shifted between different formats, some of them quite frequently (e.g., Nils' and Doris' lessons). Some shifts were clearly planned beforehand, but others were decided on the spot. For instance, Íris observed many student pairs struggling with the task of constructing a probability tree. She reacted by shifting from pair work to a brief segment of direct instruction to the whole class about probability trees from the whiteboard. Afterwards, students resumed their pair work.

The results of the content analysis of these cognitively activating lessons can be summarised in two main points. First, the instructional formats between the lessons varied greatly, with some lessons prioritising time on students' group-work, others on whole-class discussion, and some on shifting between formats. There was not a single shared characteristic of instructional format – but direct instruction was present in all lessons in very brief intervals, and where individual work was present, it was in short sprints before moving to another instructional format. Second, the teachers tended to select specific tasks for all students to work on at the same time, commonly in groups of two to four, or in a whole-class discussion. However, these tasks were not always mathematically rich in nature, meaning it was the teachers' implementation of the tasks rather than the tasks themselves that contributed to the high cognitive activation of the lessons. The themes on the teacher-student interactions further illuminate how the teachers implemented the tasks so that cognitive activation could be considered high.

4.3.2 Theme 1: Frequent shifts in types of interactions

Some lessons shifted frequently between instructional formats, but the types of interactions that teachers engaged in with students shifted frequently as well. The coded interactions from the teacher included many different types of interactions, such as explanations, stating a purpose, giving feedback, making connections, and "prompting" students. While a considerable amount of lesson time was used for explanations, they were commonly interspersed with student input through prompting. The prompts were different in nature, such as to state a solution or result (e.g., Nils: "Is 27 a prime number by the way?"), or to share understandings with the class (e.g., Doris: "What have you tried to do you girls?"). While some prompts directly requested students' explanation with how or why questions (e.g., Sabrine: "Why do we take 51?"), a common interaction sequence was identified as prompt-feedback-explanation with student contributions in between. Prompt and feedback often looped before shifting to explanation. Often the explanation was done by the teacher, or it was supported by student input. In **Figure 11** is an example of a prompt-feedback-explanation sequence from Daniel's lesson during work on the age task of Jörgen and his father. Below that is another example from soon after where the prompt and feedback parts loop with the teacher requesting multiple student contributions and revoicing before redirecting the evaluation to the students.

There were also cases of the explanation part being done by students, such as in Ída's, Sabrine's, and Nils' lessons. In Ída's lesson, this took place at the end in the game of Twenty-four (further illustrated in chapter 4.1.1). In Sabrine's lesson, a student explained on the whiteboard to the whole class her group's solution path to the "flowerbed pattern" task (see **Figure 10**). The following excerpt is from Nils' lesson where he has

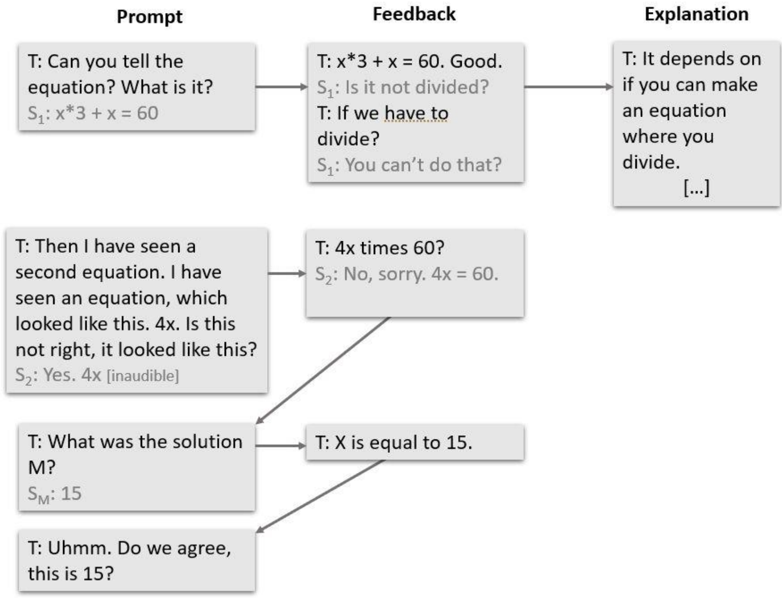


Figure 11. Examples of interactions shifting from prompting to feedback and explanation and looping between prompt and feedback.

monitored students trying to identify fractions from diagrams with coloured areas. The diagram (on **Figure 10**, by Nils on the left) was visible to all the students through a projector. Now Nils seeks explanations from students:

> Nils: Ok. What I'm wondering about... is this one. Both what the answer is and how you were thinking. What is... how did you do it H?
> H: the answer is five eighteenths.
> Nils: Five eighteenths. Mhm, because... I have not checked the answer, but it is probably correct. But it is divided into... so it is actually parts, and there are 18 parts? Yes, there are 1, 2, 3, 4, 5 parts that are coloured, and then there are 18 such triangles in total, like parts of the whole. I'm just wondering... one way to do it is to count how many parts there are. Count like, 1, 2, 3, 4, 5, 6, [...] 15, 16, 17, 18. Did anyone do it in an easier way? S, how did you do it?
> S: I took 3 times 3, because that's the quadrilateral that contains quadrilaterals.
> Nils: Where did you get... where did you see three?
> S: If you take each side, there are small quadrilaterals in there.
> Nils: Yes, those quadrilaterals there?
> S: Yes.
> Nils: There are three... three over there.
> S: Yes, so it will be three times three, and then I multiply it with two since all the quadrilaterals are divided into two.

> Nils: That's really clever. That's clever. Then you were able to see that there are kind of like small squares here, three there, three there, three there. Three, six, nine, and then since each of them were divided into to two, there is 18 parts. Very good.

By eliciting the student's explanation, Nils has made a detail explicit: how to utilise a pattern-seeking strategy to more efficiently determine the denominator of the fraction represented in the fraction diagram. This has contributed to the cognitive activation potential in the lesson, more than the task alone would. Within the interaction, he prompts the student to assert his strategy and explain it, provides the student with feedback, and then takes up the students' idea to explain it further to the whole class with reference to the fraction diagram.

Rapid shifts between types of interactions, such as prompt, feedback, and explanation, were common in the lessons, which instigated a feeling of diversity and variation. As illustrated in the example from Nils' lesson, this contributed to the cognitive activation potential of the lesson by inviting students to explain and assert rather than accept. If the interactions had solely involved explanations made by the teacher, without prompting students to actively contribute and feedback provided on their contributions, the active engagement would have been less, and thus a lower potential for cognitive activation.

4.3.3 Theme 2: Use of formative feedback and explicit student roles

As mentioned in the previous theme, a particular type of interaction commonly observed in the lessons was feedback to students. Generally, the teachers' feedback ranged from superficial (e.g., "Good" or "Correct"), to giving hints (e.g, Daniel: "Your hint right now, it is that in your equation, I want you to think of an x") or explicit formative feedback drawing on comparisons. This showed teachers giving students opportunities to explain their current understandings, which allowed them to check for students' comprehension. By uptake of students' ideas in whole-class discussions (such as Daniel revoicing their contributions in **Figure 11** and in the excerpt from Nils), the students were given agency in how the lesson progressed. For instance, in the beginning of Nadia's revision lesson, she asked the whole class to show a thumb up, sideways, or down regarding their confidence in their division competency.

> Nadia: Show me, like, if we say that, yeah, we'll use the thumb. Division, to divide. Easy, medium or difficult. Show me with the thumb. There are actually quite a few who have this one [Medium]. Some are here [Easy], that's good. Very many have this here [Medium], that they are insecure. And then I thought, that we should learn this for real in this lesson here; once and for all learn division.

She then provided each student with a small whiteboard. She read out loud a word problem about spreading a bouquet of 18 roses equally between a bedroom, a living room, and a kitchen. Each student solved the problem individually on the small whiteboard and showed the teacher to receive immediate feedback from Nadia: "Okay, the last four are here. Ok, now I want to check the rest. Right, right, right... Ask D, right, right. Perfect. Great. Great, great." The process was then repeated for another (spoken) word problem about sharing 63 chocolates between seven people.

Further examples of feedback can be taken from Íris' and Sandra's lessons. Íris provided feedback to students while they were drawing probability trees with short comments such as: "You forget... you are forgetting two branches there", and: "There you get a bit tight with space, well, it's all good". Sandra frequently got into long feedback exchanges with student pairs working on the tangram task. In the following example, students were perplexed by what fraction a tangram piece (marked E in **Figure 10**) represented in connection to the whole. Sandra suggested comparing it to another piece that the students have identified as one quarter:

> Sandra: If you think that you can actually split that and then you might see that, well, that is half of it. And how do you write it in fractional form then? Half of a quarter?
> E: Do you write like, a four the eighth? Or?
> Sandra: Yes, you think right, but it cannot be four. If we think like this. I think it's good to draw a circle like this. Then you know, you know that it's a quarter. If I'm going to split these into half as large pieces, what do we call them then? You said eighth, but you said four eighths. But... is not that... we have 1, 2, 3, 4, 5, 6, 7, 8 pieces in that circle now. And one of them, we write as one eighth. And how much is that half of, a quarter? Because I wonder where you got the eight from. How you thought...
> E: If you double four.
> Sandra: Yes.
> E: With two, there will be eight. And if you divide ... I know, a quarter, that is, four pieces, then it should be eight.
> Sandra: Right? So it will be eighths, that's right. And then you can actually see that each like this is an eighth, one of eight pieces.
> E: Mhm.
> Sandra: Did you get proof then that half of a quarter is an eighth?
> E: Mhm.
> Sandra: Or, are you not convinced?

Sandra has weaved together an explanation of how to approach the task with feedback on what the students have already demonstrated on a quarter. This interaction then developed into a discussion about a circle representation of half of a quarter and then its connection to one-sixteenth. The exchange is mostly directed by the teacher, while the students' implicit role is to express their current understanding and receive feedback.

Some teachers, such as Nadia and Sabrine, used explicit student roles for group-work. Nadia assigned pair-work toward the end of her lesson where students assumed the role of professor and secretary. The student in the role of professor was to explain, verbally only, how to solve a division task. The student in the role of secretary wrote out the calculations according to the verbal explanations. Every student pair did at least one task in each role. Nadia also referred to the "Four B's" for individual work, which represented: Brain, Book, Buddy, Boss. This referred the sequence of ways in which students should turn in attempting to solve a task – use the brain first, then check the book, and ask a buddy before raising their hand for assistance from the boss (i.e., the teacher). The professor and secretary roles aimed to facilitate student-pair interactions, while the Four B's aimed to make students more independent in their learning but also to encourage student interaction before asking for help from the teacher. For another example of explicit roles, Sabrine used what she called a "TEAM-model" for group-work where each letter represented a certain student role within the group. T made sure every student in the group understood the task, E was to read the task out loud for the group, A was responsible for presenting the group's solution, and M assumed the role of secretary, i.e., writing the group's calculations.

The theme on use of formative feedback and explicit student roles and its connection to cognitive activation can be summarised in two main points. The feedback teachers provided to their students supported cognitive activation by implicitly putting students in the role of explaining their current understanding, thus facilitating their cognitive activity and, such as in Sandra's example, providing feedback with reference to mathematically rich visual models. The explicit roles given to students provided them with opportunity to give feedback to each other and activates them toward making sense of the mathematical procedures they carried out. Granted, the focus was at times on procedures – but the teachers also strived to make connections to concepts and key ideas.

4.3.4 Theme 3: Connection-making within mathematics and to non-mathematical experience

Many teachers engaged in "connection-making" interactions. What exemplified these interactions was both connecting the mathematical content to student's previous experiences or daily lives, and to draw connections within mathematics, such as between two concepts or between concepts and methods. One can interpret that the aim of these interactions was to move students toward relational understanding of mathematical concepts.

Connection-making interactions were observed where teachers motivated students to understand a purpose of mathematical topics. Teachers did this in different ways. Some provided motivation by connecting to non-mathematical experience while others did so with a pure mathematics connection. For instance, when beginning a whole-class discussion, Ída motivated using letters in algebraic expressions by suggesting its application in calculating revenue for the local ski resort:

> Ída: Now we are going to look, because we are in the algebra chapter, and when we learn algebra when we advance through school then they start to come in, you see, letters.
> S: Letters…
> Ída: Yes. X and Y and A and B and such, something which signifies as unknown quantity. Something that we do not know. And even though you may not see the purpose of it, then it has an incredible… it shows up in many places in business, where something is being estimated. When they estimate what the revenue will be this year, in the ski resort, how many do you think will come? It costs maybe 1200 kr. for adults and then you sometimes just think, the revenue is 1200 times X. And X would be what then, do you think?
> S: An unknown number.
> Ída: Yes, an unknown number. But what do you think this X represents?
> S: The people that would come.
> Ída: The people that would come. But we do not know how many come to the ski resort, do we? So this is used, and it shows up in incredibly many places.

In this interaction, Ída has activated students' prior knowledge on the application of calculating revenue by multiplication of two quantities but connected to the abstraction of using a letter to symbolise a variable for one unknown quantity. Following this interaction, students received small whiteboards to collaboratively come up with algebraic expressions for a word problem on planting trees, shared with the whole class using a projector.

Nils took a different approach at the beginning of his lesson. He motivated students to understand the purpose of fractions by connecting to decimals. He started by saying:

> Nils: We start with fractions now, you've had it a bit in elementary school, and then we start now at lower secondary. And it is something we continue with throughout all mathematics. Calculation with fractions. If, or when, you become proficient with fractions, you will be able to do things, with, in math, that… you almost didn't even think was possible, it is very useful, to be able to do calculations with fractions. And, when you understand it… have practiced and understood it… it is not that difficult either.

Nils then started to write out the infinite decimal for one-seventh (0.142857...) and asked for a far easier way to write that number. He was first suggested with 1/6, and then with the fraction representation 1/7. He continued:

> Nils: One of the benefits with fractions, is that, you can write numbers in a simple way… that you could not… like there is no, like there is no, like… it is the easiest way we use to write many numbers. Some numbers are pretty easy to write. For example, 0.5. That's not very hard to write. But how would we write it as a fraction? S?
> S: One half.
> Nils: Yes, one half. And then you can, just like you say, you can write it as five tenths. And… you can write in several ways. You could write it as a fraction. One half and two fourths, if you were to say one more, what would it be then?
> S: Three sixths.
> Nils: Three sixths. And I can also put on another thing. Four eights. So, this is several different fractions, that are all one half. And I will try to show you that now.

In contrast to Ída's approach to connect to non-mathematical experience, Nils introduced his topic by connecting to another mathematical concept. One can interpret this as a mathematically rich motivation – it provides opportunity for students to understand why fractions are oftentimes a more practical representation of numbers than decimals. While Nils did not expect students to figure this reason out entirely on their own, he made this detail explicit as a motivation for understanding the purpose of the lesson's topic. Following this interaction, he assigned students with specific tasks where they determined the numerator and denominator from visual models (fraction diagrams), and then evaluated and compared fractions.

All the lessons showed an emphasis on student understanding. Although the focus was sometimes on procedures, this focus was not without reflection on the procedures or their connections to other ideas. There were examples of meta-cognitive strategies where students reflected and made connections to further learning. For instance, after the professor and secretary task, Nadia's lesson concluded with students being asked to write two things: one thing they learned during the lesson, and one thing they still wonder about. Nadia asked some students to share with the class:

> Nadia: I want three people to share something they learned during this lesson. Tell me everyone. And then everyone has to listen. C?
> C: I got better at division, I wonder how I divide with commas
> Nadia: Divide with commas. Great. I'll write it like that. What, is there anyone else who can say what they have learned during this lesson or something they are wondering about? M?

M: I have learned to divide.
Nadia: Have you learned to divide? Good, that was what we were going to learn this lesson. Is there anything you are still wondering about? Which you have not learnt? Divide with commas?
M: Yes.
Nadia: How about you S?
S: I have become better at that new division method there.
Nadia: This one?
S: Yes.
Nadia: What division method have you learned from before?
S: One of those box ones that used 1000 years.
Nadia: That one?
S: Yes.
Nadia: You learnt that before? How many people use this one here? In the class. Yes, a couple, I could see that you did it in the start A and R, didn't you also write it like that?
R: No.
Nadia: Oh well, no no. But good. Any questions?
S: I also want to be better at turning percentages into fractions, and such.
Nadia: Percentages to fractions and such, yes, then we have fractions. Great, H?
H: I want to be better at calculating percentages, like I know 50, 25, 100. But like I struggle with 23 for example then, or yeah.
Nadia: How many people in the class wants, or who find it difficult with fractions and decimal numbers, and percentages, and such? How many people think it's a bit tricky? Then we will talk about that when we are done with the four calculation operations. Isn't that fine?
E: Yes.

In this way, the connections that students themselves had made in their reflections were made explicit.

Connection-making is not explicitly an aspect of cognitive activation. However, the way it manifested in these lessons raises questions of their association. The way Ída and Nils introduced students to the mathematical topics in different ways highlights the way they made details of their lesson's content explicit for students to see connections and develop understanding in the tasks that followed. The reflection task at the end of Nadia's lesson was an example that encouraged students to think about their thinking, i.e., develop their metacognition. This task also invited connections to other ideas for further learning. The emphasis in the lessons on making connections, understanding, and promoting multiple solution paths is exemplified in Nils' words: "There is no right or wrong way. The important thing is that you can explain it mathematically."

4.4 Summary

The project's findings have enriched knowledge about cognitive activation in mathematics lessons in Iceland from different perspectives and about cognitively activating lessons in a Nordic context. Observation scores on cognitive activation in Iceland were mostly on the low end, and so were the mathematical tasks. The potential for cognitive activation can therefore be described as mostly low. Two out of ten teachers had segments that scored at the highest level in IC and CD. However, student perceptions of cognitive activation were mostly on the high-end, although often lower than other items in the survey. The connection between observation scores and student perception as measures of cognitive activation in Iceland was found to be weak, to some extent explained by a high variation in student perceptions within classrooms and a low variation between them. In the analysis of eight Nordic mathematics lessons with outstanding scores in cognitive activation, the instructional formats were varied, with lessons often focused on group-work and whole-class discussions. Some brief direct instruction was identified in all lessons. Where individual work was found it was in short bursts before moving to a different format. The teacher-student interactions were characterised by frequent shifts in types of interactions, with types ranging from prompting students, providing feedback, explanations, stating a purpose, and making connections. Use of formative feedback was common and student roles in the classroom were often made explicit. Teachers commonly engaged in connection-making, both within mathematics and to non-mathematical experience, which can be interpreted as moving students toward a relational understanding of mathematics. Examples of teacher-student interactions illustrated how these themes contributed or connected to the high cognitive activation in the lessons.

5 Discussion

A presentation of this dissertation's contribution to the educational research field is found in this chapter. The overarching aim of this doctoral dissertation was to develop a deeper understanding of the teaching quality dimension of cognitive activation in mathematics in Iceland and in a Nordic context. The discussion is outlined by what the findings put forward in an empirical, theoretical, and methodological contribution. Lastly, I offer some final reflections and concluding remarks.

5.1 Empirical contribution

The aims of the dissertation had a strong empirical focus, so perhaps the most evident value is in its empirical contribution. This contribution can be identified throughout the findings: with empirical descriptions of different levels of cognitive activation in Iceland through systematic classroom observation and analysis of tasks, the quantitative analysis revealing a weak connection between student perceptions and observation scores as measures of cognitive activation, and qualitative descriptions of interactions and instructional formats in highly cognitively activating lessons in a Nordic context.

The findings showed that the potential for cognitive activation in the observed lower secondary mathematics classrooms in Iceland was mostly low. The reasons for these low scores were largely due to teachers reducing the challenge of the tasks. This often occurred within a teaching format where teachers assisted individual students working at their own pace according to a pre-determined plan with numbered exercises to work on. The emphasis on individual work is consistent with previous findings on mathematics teaching in Iceland, where individual work in textbooks has been found to be the dominant format (Gunnarsdóttir & Pálsdóttir, 2015; Jónsdóttir et al., 2014; Savola, 2010; Þórðardóttir & Hermannsson, 2012). What these findings contribute to the empirical understandings of mathematics teaching in Iceland is the qualitatively rich descriptions of cognitive activation as systematically rated at different levels. As such, they further solidify previous findings and suggest possible pathways forward to improve cognitive activation in mathematics teaching in Iceland. With high turnover in the teaching profession and a large number of teachers in Iceland approaching retirement, it is imperative to find ways to support both practicing and pre-service teachers in developing the necessary knowledge, skills, and attitudes to teach mathematics that gives students opportunities to productively struggle and develop understanding (Gíslason & Gísladóttir, 2021; Hreinsdóttir & Diego, 2019; Tekkumru-Kisa et al., 2020; Wilhelm, 2014). Rich opportunities exist to equip future teachers with cognitive activation strategies, so that they plan lessons where students have opportunity to explain their thinking and to work collaboratively on mathematically rich tasks. This can be the object of further development and study with respect to these findings.

A related empirical contribution is in the tasks identified as doing mathematics (TAG-4). Four such tasks were identified from four separate teachers: T_1 and T_2 from group A, and T_8 and T_{10} from group C. In the cases of T_1 and T_2, the tasks were implemented in a way that resulted in high observation scores. However, the tasks from T_8 and T_{10} were enacted within a classroom where the social norm is that students work individually at their own pace (Yackel & Cobb, 1996). Students therefore encountered these tasks as "solo performers". This did not lead to fruitful mathematical discussions or collective problem solving. It is unclear to what extent students in these individual work lessons engaged productively with these problems. These empirical results may be useful examples to consider, for either pre-service mathematics teachers or for professional development, in exploring ways of implementing different mathematical tasks. The distribution of tasks across the four TAG levels in the findings is similar to previous findings of tasks in upper secondary level textbooks in Iceland, indicating a strong focus on procedural competency and less on connections to mathematical concepts (Sigurjónsson, 2014).

Another empirical contribution is in the student perception data. The evidence shows that variation in student perceptions of teaching is oftentimes greater within classrooms than between them. This creates difficulties for interpretation and inferences, as it indicates disagreement among students on what goes on in the classroom (Sandilos et al., 2019). A part of this can possibly be explained by the teacher acting differently toward different students. There is more work to be done in disentangling the somewhat paradoxical empirical results between student perceptions and systematic observer ratings, such as those seen in the GTI study (OECD, 2020). In Paper II, a certain "cognitive activation paradox" is proposed, between the observed cognitive activation in lessons and what students perceive in their cognitive engagement. Caution has been advised for using student perceptions alone to measure cognitive activation, as with other constructs (Kuhfeld, 2017; Phillips et al., 2021). However, student voices can be vital for exploring other aspects of teaching and learning. This may include classroom management or students' experiences of a positive classroom climate for learning, which may be more difficult for observers to rate directly from observing a limited number of lessons. The findings contribute to empirical understandings of student perceptions of cognitive activation. Further research is suggested in chapter 5.3 on methodological contribution.

A novel contribution is made to understandings of cognitive activation in Nordic mathematics teaching in the rich empirical descriptions of teacher-student interactions and instructional formats in cognitively activating mathematics lessons. A conclusion from these findings is that there is no "blueprint" or "recipe" for a cognitively activating lesson. However, I can suggest possible directions for teachers and prospective teachers to develop their cognitive activation. These directions may involve engaging in connection-making between mathematical concepts and outside the mathematical world, as well as to provide formative feedback in interactions with students. Teachers may also benefit from reflecting on the types of interactions they make with their students and what instructional formats they prioritise.

It may be noted that these outstanding lessons are likely far from typical Nordic mathematics lessons. Previous studies on Nordic mathematics teaching have indicated a focus on procedural fluency, which may be interpreted as low potential for cognitive activation (Bergem & Pepin, 2013; Boesen et al., 2014; Kelly et al., 2013; Stovner & Klette, 2022; Tengberg et al., 2021). Crucially, these findings are developed from observations of teaching in a Nordic cultural and social context. The contribution may therefore be limited to understanding teaching practices within a similar context, such as where individual seatwork is the most common instructional format.

One of the mentioned limitations of the study was the limited information about the student groups. A potential for further study is considering the effect of cognitive activation strategies for different groups of students. Prior research has indicated a connection to increased student enjoyment of mathematics, especially among girls (Cantley et al., 2017; Lazarides & Buchholz, 2019). Further inquiry is warranted into how cognitive activation teaching strategies affect high-achieving students, and whether their perceptions of teaching practices with high cognitive activation may be different from more disadvantaged students. This would essentially move the focus to the student use of the cognitive activation potential they are offered, according to the offer-use model (Helmke, 2015; Weingartner, 2021). Such results may inform to what extent characteristics of different groups of students can explain the high variation in student perceptions of cognitive activation within classrooms.

5.2 Theoretical contribution

The theoretical contribution is mostly indirect and with potential to develop further in future research. To a large extent, this stems from the doctoral project's aims that were fundamentally more geared toward an empirical and methodological contribution rather than theory development. For example, one aim was to describe the connection between student perceptions and observation scores as measures of cognitive activation – the aim was not to develop an explanatory framework for the discrepancy that was found. For such work, more data and more time would be required. This sub-chapter will therefore not list a great contribution to theory. Rather, it will provide suggestions for future research and potentials for theory development with reference to the findings and the theoretical framing that the project is built on.

One of the criticisms of cognitive activation is that it is not a sufficiently well-defined concept. The definition built on in this dissertation is the interpretation of Praetorius and Charalambous (2018). They build on Klieme (et al., 2001) who first put forth the concept with a theoretical foundation in a text on educational philosophy by Diederich and Tenorth (1997). Throughout the years, the specific components of cognitive activation as a theoretical construct appear to shift slightly in the literature even though the fundamental vision of working toward student understanding remains intact. For instance, the component *errors as opportunities* from the original conceptualisation is not explicitly included in the Praetorius and Charalambous (2018) interpretation. However, it was included in the way cognitive activation was measured in the COACTIV project in a survey item about whether "the teacher helps us to learn from mistakes we have made" (Kunter & Voss, 2013; Neubrand et al., 2013).

Another proposed component of cognitive activation not included in Praetorius and Charalambous' definition is activating prior knowledge (Klieme et al., 2006; von Kotzebue et al., 2020). Keeping these separate is consistent with the way teaching quality is conceptualised in PLATO, where the element of Connections to Prior Knowledge is listed under the instructional domain Representation and Use of Content – separate from IC and CD, which measure cognitive activation (or Disciplinary Demand as the PLATO instructional domain is named, see chapter 3.5.1; Bell et al., 2019; Grossman, 2015). However, one of the themes on the teacher-student interactions in the outstanding lessons in cognitive activation was about the connection-making that the teachers engaged in. This may suggest that activating prior knowledge could potentially be considered a component of cognitive activation rather than content representation. Future research may examine further the theoretical grounding of the "connections" aspect of teaching. The most pressing question is perhaps not under which hat it should be placed – rather, what implications it makes to theorise it in one way and not in another. The findings on connection-making may be a relevant contribution for theoretical work that examines and addresses the weight of the component of activating prior knowledge for cognitive activation.

Relatedly, the theme on formative feedback raises some theoretical questions. Formative feedback was quite apparent in some of the outstanding Nordic lessons in cognitive activation. The extent and meaning of this connection are worthy of further study, possibly in relation to how cognitive activation and scaffolding can be intertwined. As currently conceptualised, some might argue that by scaffolding tasks, teachers will lower the cognitive activation potential. Finding synergies between these concepts to produce strategies for teachers to scaffold tasks in a way that does not diminish students' productive struggle can be the object of future work. Regular, formative feedback as observed in the findings may be a starting point in developing such strategies. The use of tools such as student whiteboards to make their thinking and work visible but non-permanent was observed in the findings (Liljedahl, 2018, 2021). This may inform developments toward "cognitively activating scaffolding".

The findings on teacher-student interactions offered some insights into that "hidden dimension" of mathematics classrooms (Bauersfeld, 1980). A frequent interaction sequence observed was what was called prompt-feedback-explanation. I would argue that the way it was manifested in Ída's, Sabrina's and Nils' lessons (as in the example in chapter 4.3.2) invited student's assert mode, i.e., explore and explain their solutions, rather than accept mode, i.e., to accept what the teacher says as truth (Mason & Johnston-Wilder, 2006), thereby exerting less teacher control than the widely known IRE pattern (Cazden, 1988). The connection-making, such as the one shown in the excerpt toward the end of Nadia's lesson, seemed to underpin conceptual understanding in a near-discussion interaction pattern (Pimm, 1994).

The findings may have implications for educational systems. With understanding as an educational goal and teachers under increasing pressure to perform, it may be necessary to offer teachers additional support and direction to develop a teaching repertoire where cognitive activation strategies are implemented with mathematically rich tasks. This may involve increased central control to some extent – at least offering guidance and suggestions for teaching, as well as possibilities for teachers to collaborate and develop their teaching competencies. However, this may be a fine line to tread, threading a delicate balance between respecting teachers' autonomy and trust in their professional judgement. The challenge is to provide adequate guidance to implement educational goals that may be demanding to work toward without specific expertise in a subject such as mathematics. I propose that for future work aiming to improve teaching quality, this will be an important issue to solve within educational systems and a topic worthy of further research and development.

Teaching quality is generally viewed as a multi-dimensional concept (Croninger et al., 2012). However, the number and scope of its dimensions remains an open question. It also remains an open question to what extent conceptualisations, or even number of dimensions, may be bound to specific cultural or social contexts. As suggested by some findings in the GTI study, the widely differential variation between contexts, from virtually no variation to considerable variation, begs the question to what extent it is meaningful or even possible to conceptualise teaching quality dimensions as global (OECD, 2020). Indeed, teaching has been recognised as a cultural activity grounded in shared knowledge yet often varied within countries – and understanding of it has come a long way since the idea of "national patterns of teaching" (Stigler & Hiebert, 1997, 1999). Conceivably, different educational goals and cultural norms should be taken into greater consideration for comparisons across contexts. Continued developments within the field towards more synergy between theoretical concepts and their context-sensitivity, which also relate to methodological developments, are to be expected.

5.3 Methodological contribution

The dissertation has made a methodological contribution with some implications for future research, specifically relating to the student perceptions and task analysis. It has combined measurement frameworks developed in the United States (PLATO, TAG, Tripod) with an interpretation from the theoretical perspective of cognitive activation developed in Germany.

In working with the Iceland classroom data, two different analyses were combined: The analysis of lesson videos using the PLATO observation system, and an analysis of the cognitive demand of tasks with the Task Analysis Guide (TAG). An observation that can be contributed is a certain discrepancy between the two rubrics when it comes to intellectual challenge dimension in PLATO (Grossman, 2015; see Appendix A) and the cognitive demand of tasks in TAG (Stein et al., 2009; see Appendix B). Even though both rubrics categorise each unit of analysis into four categories, i.e., a lesson segment on a scale from 1 to 4 and a task on a scale from 1 to 4, these scales do not correspond to each other on each level. Perhaps most notably, a task categorised as Procedures without connections (TAG-2) would be considered activity at the 3-level when rated for intellectual challenge in PLATO. In other words, PLATO does not distinguish between mathematical tasks being procedural with or without connections to concepts, as is a key distinguishing component in TAG. In some ways, it is problematic to use two instruments measuring a similar construct in four categories where the distinction between the middle two categories diverges in this way. However, in the findings it has been shown that it can be useful to analyse the tasks separately, specifically in a context where individual work in textbooks is so prevalent (the information contained in the volume of tasks across classrooms is a prominent example). An implication is that future research could explore synergies across different measurements, such as rubrics analysing the cognitive potential of tasks, and frameworks analysing the enacted cognitive activation during lessons.

The student survey findings contribute to a critical discussion of how to interpret student perceptions and, more specifically, the wording of some Tripod items. In the REAS1 item in the reasoning subscale ("My teacher asks questions to be sure we are following along when s/he is teaching"), it is entirely possible that many teachers did ask their students such questions yet proceeded to assist them in ways that diminished their productive struggle, resulting in high student ratings but low observation scores. REAS1 is in less agreement with other related items, which matches previous findings in a large sample (where it had the highest mean score across all items in the survey, and the second lowest intra-class correlation; see item labelled CHAL1 in Schweig, 2014). In the discourse subscale, a possible explanation for T_2's average student rating compared to the observation evidence is the fact that even though she had segments with a high level of student discourse, she also had lesson segments where they work in silence, such as each student watching instructional videos with headphones on. The observation shows that she does not "always" offer students opportunities for content-related talk. How the students interpret "always" in the items is a question worth exploring. Do they take it to mean constantly in every single lesson, at least once in every lesson, or does once a week suffice? Strengthening understandings of how student respondents may interpret item choices in different contexts with rigorous pilot testing or student interviews will allow stronger inferences to be drawn from student survey data. The findings suggest that among the dimensions of teaching measured by classroom observations, the cognitive activation dimension has potential for developments in synthesising instruments, such as the Tripod survey and PLATO, to measure its different aspects more accurately.

The results suggest a discrepancy between student perceptions and classroom observations as measures of cognitive activation. To what extent this discrepancy is actual and to what extent it may be due to students' different interpretations remains unclear. Nevertheless, the results warrant some caution in using student perceptions, at least with the Tripod instrument, to evaluate cognitive activation. Similarly, previous studies have cautioned against using student perception surveys for high-stakes decisions (Phillips et al., 2021; Wallace et al., 2016). The findings give a reason to doubt that cognitively activating instruction, as theorised and measured by systematic observation, is directly connected to students experiencing instruction as cognitively activating. Further inquiry into connections between dimensions of teaching quality as observed by researchers and rated by students will support more reliable and valid measures for improving teaching practice.

The ordered grouping of teachers according to the observation evidence was an approach that prioritised the maximum scores rather than using averages. It is worth commenting that an extension of this approach to other dimensions, within PLATO or other observation systems, is possible – but this must be carefully planned with respect to the data at hand. It may be differently suitable to different distributions of teaching quality dimensions. For instance, in a dimension with a heavily skewed distribution, all teachers may at some point receive a score at the 4-level, placing them all in group A if the same procedure is followed. This approach is perhaps most suitable in distributions where high-level scores are relatively rare. Further, observation systems do differ in how many levels they include in the scales (Bell et al., 2019). PLATO has four levels in every dimension but applying the procedure to a dimension with more levels may result in the number of groups to be greater, depending on how the procedure would be adapted.

Another conclusion from the findings on the Nordic lessons is that the implementation of tasks weighed more heavily than the selection of tasks in deeming the lessons cognitively activating. Prior studies have rated cognitive activation solely based on the selection of tasks (e.g., Neubrand et al., 2013). Although task selection may be a reliable indicator in some contexts, the findings illustrate cases where the teachers have selected tasks likely deemed low in cognitive activation potential but implemented them in such a way that the students must explain, discuss, and reason for their solutions. An example of this is in Nadia's lessons where students assumed the roles of professor and secretary in explaining to each other their solution process to otherwise rather standard procedural division tasks. This finding highlights some shortcomings of solely relying on task analysis to estimate the cognitive activation potential of a lesson. The task analysis results from the Iceland data may further suggest that the cultural context, i.e., the dominance of the individual seatwork format, possibly plays a part in whether task demand alone can be a good indicator of cognitive activation.

5.4 Concluding remarks

This doctoral research project set out with the overarching aim to develop a deeper understanding of cognitive activation in mathematics teaching in Iceland and the Nordic countries. At this point, after designing the project, participating in data collection, analysing data, and writing up results, my impression is that the aim has been achieved. As discussed, and argued for in this chapter, the findings have implications for local policy and practice, both in solidifying previous findings on mathematics teaching in Iceland and suggesting pathways forwards. They also contribute to methodology and raise some questions for theory development. However, there is a long way to go to understand Nordic teaching more fully and to develop the way mathematics is taught and what teaching and learning mathematics is understood to be. For this concluding chapter, I would like to reflect on these issues.

Teaching is sometimes regarded as both an art and a science. I have previously pondered on what constitutes "the art of being a mathematics teacher", in the context of what should be directly explained to students and what should be left for them to discover. In this dissertation, my theoretical view of teaching has leaned more toward teaching as a science, with aspects, features, or dimensions that can be of varying and measurable quality.

While I can align myself somewhat with the view of teaching as both an art and a science, I would argue that more fundamentally, teaching mathematics is a human endeavour within a certain cultural and systemic context. The Icelandic word for education, "menntun", captures this perspective: to "humanise", or become human. More specifically, this also endorses Pólya's notion that doing mathematics is fundamentally a human endeavour (1981). Mathematics does not just entail but is essentially about imagination, play, formulating hypothesis, failing, making mistakes, and learning something along the way. The role of teachers in mathematics is to guide students in this endeavour, with all its messiness and complexity, with the aim of developing some understanding – learning not just procedures, but how they work and connect to big ideas and abstract concepts.

Cognitive activation, as an overarching concept of some aspects of teaching quality, explains to what extent the teaching is geared toward understanding, and in my view is also aimed at to what extent the process is humanistic. This is meant in the sense that when cognitive activation is low, students are not engaged in playing with mathematics, trying to make sense of its concepts. When cognitive activation is high, students are engaged in communicating their thoughts, ideas, and understandings of varying degrees of complexity. Two students can both be cognitively activated even though they are currently at different levels of understanding of the concepts at hand. What matters is that the act of participating in a mathematics classroom involves that this curiosity-fuelled human act of doing mathematics is preserved. This can be realised in several ways: by playing games, engaging in a discussion, explaining an idea, or communicating a current hypothesis or state of understanding. That is a human endeavour. However, many familiar classroom routines do not subscribe to this view of mathematics teaching and learning. For instance, this vision is not realised by memorising facts, routinising a procedure by drill exercises, doing only odd-numbered exercises because the lesson plan says so, and being scolded if behind schedule. That, I argue, is not a humane way of doing mathematics. I do acknowledge that fluency in executing procedures can be useful. But usefulness has typically been a side-product of mathematics rather than its central goal. Teachers currently engrained in some of these routines, that I do sympathise with, may not be aware of other ways or their own potential to develop new and more humanistic routines. Perhaps that rarely happens spontaneously.

A common use of lesson time in mathematics in Iceland is students' individual work on numbered exercises listed in a plan from the teacher. In a classroom culture where doing (mostly procedural) exercises according to a pre-determined plan and students being assigned homework if "not on plan" is a social norm, I argue that the incentive for students to engage in joint sense-making and active knowledge construction is worked against. For example, consider a student who makes good progress with the plan in class. The student may risk losing free time to homework if time is spent explaining to other students. This also raises questions on equal opportunities to learn, as high-achieving students may be more likely to receive assistance at home, while low-achieving students may be less likely to have the same support outside the school and are therefore more likely to lag behind (for instance, this may apply to students with small or no extended family in the area, such as children of immigrants or other marginalised social groups). Furthermore, I raise the question of what teachers are capable of in the situation that they are given. In an understaffed school system intended to implement inclusive schooling, teachers of mathematics and other specialised subjects may be left with little support to face student groups with a very wide range of competencies. For teachers in this situation, it may feel overwhelming to plan ambitious teaching where all students are to participate in the same task at the same time. Shifting from a pre-determined exercise plan to a single challenging task may raise teacher's concerns about different student groups. They may fear low-achieving students may feel "left behind", or risk high-achieving students feeling that they "waste time" by participating. Possibly, some teachers experience that they are constrained by this working environment and that they lack the time to prepare ambitious lessons with rich opportunities for cognitive activation (Teig et al., 2019). Teachers' perspectives on these issues and ways forward to develop cognitive activation practices under these constraints may be the subject of further study. The questions going forward, I argue, have as much to do with working with teachers as it has to do with developing a system in which a passion for high quality teaching can thrive. And, as indicated in this dissertation's findings, simple uniform solutions or "recipes" are unlikely to be the answer.

In my view, opportunities for student reflections, mathematically rich discussions and critically engaging with challenging tasks are not important because they work in some sense – these opportunities are important because they matter for how mathematics is socially constructed by experiences that shape people's understandings of what it means to do mathematics. Mathematics teaching dominated by individual seatwork on pre-determined numbered textbook exercises in a linear fashion across many lessons does not adequately construct mathematics as a human endeavour. Combined through years and decades, these experiences contribute to creating a culture of mathematics as an exclusionary subject at best, or worse, a subject feared and despised. Within the frame of teaching as a teacher's actions with varying degrees of quality, and teachers as professionals capable of developing their profession toward improved practices, it does not suffice to refer to tradition when it comes to preparing and organising lessons. Teachers must be given the necessary time to systematically reflect on their lessons and develop their teaching.

High quality teaching is not pursued with vision and passion alone. Teaching is the profession of teachers. Teachers belong to a community of professionals within the school and in the wider community. Being an ambitious teacher requires working along with other professionals to ensure that the vision of education that the community agrees on is realised. It also requires each teacher to have pedagogical knowledge and skills, be it content-specific knowledge or more general pedagogical skills, as well as a positive view of the subject. Mathematics is known for people's negative views of it, even among prospective teachers (Gíslason & Gísladóttir, 2021). Some of those prospective teachers eventually come to teach mathematics. Even with only good intentions, it can be a challenge to hinder students reproducing these negative views and passing them on to another generation. Previous studies have suggested that high cognitive activation, as most commonly enacted by teachers with strong pedagogical content knowledge, may strengthen students' positive views of mathematics and student achievement. It is my sincere hope that with this doctoral dissertation, I have offered at least one brick unto the bridge toward understanding what encompasses such teaching practices and how researchers may identify and measure them – and thus, contributed to supporting more teachers in the future to engage in teaching practices that make more students enjoy and succeed in a productive struggle with mathematics.

References

Andersson, E., & Sørvik, G. O. (2013). Reality lost? Re-use of qualitative data in classroom video studies. *Forum: Qualitative social research*, *14*(3).

Bauersfeld, H. (1980). Hidden dimensions in the so-called reality of a mathematics classroom. *Educational studies in mathematics*, *11*(1), 23–41. https://doi.org/10.1007/BF00369158

Bauersfeld, H. (1994). Theoretical perspectives on interaction in the mathematics classroom. In R. Biehler, R. W. Scholz, R. Strässer, & B. Winkelmann (Eds.), *Didactics of mathematics as a scientific discipline* (pp. 133–146). Kluwer.

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

Baumert, J., Kunter, M., Blum, W., Klusmann, U., Krauss, S., & Neubrand, M. (2013). Professional competence of teachers, cognitively activating instruction, and the development of students' mathematical literacy (COACTIV): A research program. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss, & M. Neubrand (Eds.), *Cognitive activation in the mathematics classroom and professional competence of teachers: Results from the COACTIV project* (pp. 1–21). Springer US. https://doi.org/10.1007/978-1-4614-5149-5_1

Bell, C. A., Dobbelaer, M. J., Klette, K., & Visscher, A. (2019). Qualities of classroom observation systems. *School effectiveness and school improvement*, *30*(1), 3–29. https://doi.org/10.1080/09243453.2018.1539014

Bell, C. A., Gitomer, D. H., McCaffrey, D. F., Hamre, B. K., Pianta, R. C., & Qi, Y. (2012). An argument approach to observation protocol validity. *Educational assessment*, *17*(2–3), 62–87. https://doi.org/10.1080/10627197.2012.715014

Bergem, O. K., & Pepin, B. (2013). Developing mathematical proficiency and democratic agency through participation – an analysis of teacher-student dialogues in a norwegian 9th grade classroom. In B. Kaur, G. Anthony, M. Ohtani, & D. Clarke (Eds.), *Student voice in mathematics classrooms around the world* (pp. 143–160). Sense.

Blikstad-Balas, M. (2017). Key challenges of using video when investigating social practices in education: contextualization, magnification, and representation. *International journal of research & method in education*, *40*(5), 511–523. https://doi.org/10.1080/1743727X.2016.1181162

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

Blossing, U., Imsen, G., & Moos, L. (2014). Nordic schools in a time of change. In U. Blossing, G. Imsen, & L. Moos (Eds.), *The Nordic education model: 'A school for all' encounters neo-liberal policy* (pp. 1–16). Springer.

Boesen, J., Helenius, O., Bergqvist, E., Bergqvist, T., Lithner, J., Palm, T., & Palmberg, B. (2014). Developing mathematical competence: From the intended to the enacted curriculum. *The journal of mathematical behavior*, *33*, 72–87. https://doi.org/https://doi.org/10.1016/j.jmathb.2013.10.001

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative research in psychology*, *3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Braun, V., & Clarke, V. (2013). *Successful qualitative research*. Sage.

Braun, V., & Clarke, V. (2022). *Thematic analysis: a practical guide*. Sage.

Buchholtz, N. (2019). Planning and conducting mixed methods studies in mathematics educational research. In G. Kaiser & N. Presmeg (Eds.), *Compendium for early career researchers in mathematics education* (pp. 131–152). Springer.

Burr, V. (2015). *Social constructionism* (3rd ed.). Routledge.

Cantley, I., Prendergast, M., & Schlindwein, F. (2017). Collaborative cognitive-activation strategies as an emancipatory force in promoting girls' interest in and enjoyment of mathematics: A cross-national case study. *International journal of educational research*, *81*, 38–51. https://doi.org/https://doi.org/10.1016/j.ijer.2016.11.004

Carroll, J. B. (1963). A model for school learning. *Teachers college record*, *64*, 723–733.

Cazden, C. B. (1988). *Classroom discourse: The language of teaching and learning*. Heinemann.

Clarke, D. (2003). International comparative studies in mathematics education. In A. J. Bishop, M. A. Clements, C. Keitel, J. Kilpatrick, & F. K. S. Leung (Eds.), *Second international handbook of mathematics education* (pp. 145–186). Kluwer.

Clarke, D. (2006). Deconstructing dichotomies: Arguing for a more inclusive approach. In D. Clarke, J. Emanuelsson, E. Jablonka, & I. A. C. Mok (Eds.), *Making connections: Comparing mathematics classrooms around the world* (pp. 215–236). Sense Publishers.

Clarke, D., & Chan, M. C. E. (2019). The use of video in classroom research: window, lens, or mirror. In L. Xu, G. Aranda, W. Widjaja, & D. Clarke (Eds.), *Video-based research in education: Cross-disciplinary perspectives* (pp. 5–18). Routledge.

Clarke, D., Emanuelsson, J., Jablonka, E., & Mok, I. A. C. (2006). The learner's perspective study and international comparisons of classroom practice. In D. Clarke, J. Emanuelsson, E. Jablonka, & I. A. C. Mok (Eds.), *Making connections: Comparing mathematics classrooms around the world* (pp. 1–22). Sense Publishers.

Clarke, D., Mesiti, C., Jablonka, E., & Shimizu, Y. (2006). Lesson structure in the USA, Germany and Japan. In *Making connections: Comparing mathematics classrooms around the world* (pp. 23–46). Sense Publishers.

Clarke, D., & Xu, L. H. (2008). Distinguishing between mathematics classrooms in Australia, China, Japan, Korea and the USA through the lens of the distribution of responsibility for knowledge generation: public oral interactivity and mathematical orality. *ZDM*, *40*(6), 963–972. https://doi.org/10.1007/s11858-008-0129-5

Cochran-Smith, M. (2003). Teaching quality matters. *Journal of teacher education*, *54*(2), 95–98. https://doi.org/10.1177/0022487102250283

Croninger, R. G., Valli, L., & Chambliss, M. J. (2012). Researching quality in teaching: Enduring and emerging challenges. *Teachers college record*, *114*(4), 1–15.

Dahl, B., & Stedøy, I. M. (2004). A Nordic community: Ideas of education and democracy in mathematics. In I. M. Stedøy (Ed.), *Mathematics education: The Nordic way* (pp. 1–10). NTNU-trykk.

Den Brok, P., Bergen, T., & Brekelmans, M. (2006). Convergence and divergence between students' and teachers' perceptions of instructional behaviour in Dutch secondary education. In D. Fisher & K. S. Khine (Eds.), *Contemporary approaches to research on learning environments: Worldviews* (pp. 125–160). World scientific. https://doi.org/10.1142/9789812774651_0006

Dewey, J. (1938). *Experience & education*. Macmillan.

Diederich, J., & Tenorth, H. E. (1997). *Theorie der Schule. Ein Studienbuch zu Geschichte, Funktionen und Gestaltung*. Cornelsen.

Doyle, W. (1983). Academic work. *Review of educational research*, *53*(2), 159–199. https://doi.org/10.3102/00346543053002159

Ekatushabe, M., Nsanganwimana, F., Muwonge, C. M., & Ssenyonga, J. (2022). The relationship between cognitive activation, self-efficacy, achievement emotions and (meta)cognitive learning strategies among ugandan biology learners. *African journal of research in mathematics, science and technology education*, *25*(3) 1–12. https://doi.org/10.1080/18117295.2021.2018867

Fauskanger, J., Naalsund, M., Nilsen, H. K., Nortvedt, G. A., Pálsdóttir, G., Portaankorva-Koivisto, P., Radišić, J., Sigurjónsson, J. Ö., Wernberg, A. (2022). Bringing Nordic mathematics education into the future – An analysis of the themes and impact of papers presented at the NORMA20 conference. In Nortvedt, G., ... Wernberg, A. (Eds.) *Proceedings of Norma 20, the ninth Nordic conference on mathematics education* (pp. 281–294).

Ferguson, R. (2010). *Student perceptions of teaching effectiveness. Discussion brief*.

Ferguson, R. F. (2012). Can student surveys measure teaching quality? *Phi Delta Kappan*, *94*(3), 24–28. https://doi.org/10.1177/003172171209400306

Furtak, E., & Shavelson, R. (2009). Guidance, conceptual understanding, and student learning: An investigation of inquiry-based teaching in the US. In T. Janík & T. Seidel (Eds.), *The power of video studies in investigating teaching and learning in the classroom* (pp. 181–206). Waxmann.

Gíslason, I., & Gísladóttir, B. (2021). Mótun viðhorfa kennaranema til stærðfræði og tengsl við árangur. *Icelandic journal of education*, *30*(2), 185–207. https://doi.org/10.24270/tuuom.2021.30.13

Gitomer, D. (2009). Crisp measurement and messy context: A clash of assumptions and metaphors—synthesis of section III. In D. Gitomer (Ed.), *Measurement issues and assessment for teaching quality* (pp. 223–233). Sage.

Givvin, K. B., Hiebert, J., Jacobs, J. K., Hollingsworth, H., & Gallimore, R. (2005). Are there national patterns of teaching? Evidence from the TIMSS 1999 video study. *Comparative education review*, *49*(3), 311–343. https://doi.org/10.1086/430260

Glasersfeld, E. von. (1995). A constructivist approach to teaching. *Constructivism in education*, 21–34. https://doi.org/10.4324/9780203052600-5

Grevholm, B. (2021). Recent Nordic research in mathematics education illustrated by examples from NORMA17. *Mathematics*, *9*(8). https://doi.org/10.3390/math9080803

Grossman, P. (2015). *Protocol for language arts teaching observations (PLATO 5.0)*. Stanford University.

Grossman, P., Loeb, S., Cohen, J., & Wyckoff, J. (2013). Measure for measure: The relationship between measures of instructional practice in middle school English language arts and teachers' value-added scores. *American journal of education*, *119*(3), 445–470. https://doi.org/10.1086/669901

Grossman, P., & McDonald, M. (2008). Back to the future: Directions for research in teaching and teacher education. *American educational research journal*, *45*(1), 184–205. https://doi.org/10.3102/0002831207312906

Gunnarsdóttir, G. H., & Pálsdóttir, G. (2015). Instructional practices in mathematics classrooms. In K. Krainer & N. Vondrová (Eds.), *Proceedings of the ninth congress of the European Society for Research in Mathematics Education* (pp. 3036–3042). ERME.

Haataja, E., Garcia Moreno-Esteva, E., Salonen, V., Laine, A., Toivanen, M., & Hannula, M. S. (2019). Teacher's visual attention when scaffolding collaborative mathematical problem solving. *Teaching and teacher education*. https://doi.org/10.1016/j.tate.2019.102877

Heath, C., Hindmarsh, J., & Luff, P. (2010). *Video in qualitative research*. Sage.

Helmke, A. (2015). *Unterrichtsqualität und Lehrerprofessionalität. Diagnose, Evaluation und Verbesserung des Unterrichts*. Kallmeyer.

Hiebert, J., Gallimore, R., Garnier, H., Givvin, K. B., Hollingsworth, H., Jacobs, J., Chui, A. M., Wearne, D., Smith, M., Kersting, N., Manaster, A., Tseng, E., Etterback, W., Manaster, C., Gonzales, P., & Stigler, J. (2003). *Teaching mathematics in seven countries: Results from the TIMSS 1999 video study*. National Center for Education Statistics.

Hiebert, J., & Grouws, D. A. (2007). The effects of classroom mathematics teaching on students' learning. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning, Vol. 1* (pp. 371–404). Information Age Publishing.

Hreinsdóttir, F. (2019). Rýnt í niðurstöður PISA um stærðfræðilæsi á Íslandi. In *PISA 2018: Helstu niðurstöður á Íslandi* (pp. 73–84). Menntamálastofnun.

Hreinsdóttir, F., & Diego, F. (2019). Stærðfræðikunnátta nema við upphaf kennaranáms. Samanburður áranna 1992 og 2014. *Netla*, 1–21. https://doi.org/10.24270/netla.2019.4

Hsieh, H.-F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative health research*, *15*(9), 1277–1288. https://doi.org/10.1177/1049732305276687

Ing, M., Webb, N. M., Franke, M. L., Turrou, A. C., Wong, J., Shin, N., & Fernandez, C. H. (2015). Student participation in elementary mathematics classrooms: the missing link between teacher practices and student achievement? *Educational studies in mathematics*, *90*(3), 341–356. https://doi.org/10.1007/s10649-015-9625-z

Jacobs, J., Garnier, H., Gallimore, R., Hollingsworth, H., Givvin, B. K., Keith, R., Kawanaka, T., Smith, M., Hiebert, D. W., Manaster, A., Etterbeek, W., Hiebert, J., & Stigler, J. (2003). *Third International Mathematics and Science Study 1999 Video Study Technical Report. Volume 1: Mathematics*. National Center for Education Statistics.

Jacobs, J. K., Hollingsworth, H., & Givvin, K. B. (2007). Video-based research made "easy": Methodological lessons learned from the TIMSS video studies. *Field methods*, *19*(3), 284–299. https://doi.org/10.1177/1525822x07302106

Jónsdóttir, A. H., Briem, E., Hreinsdóttir, F., Þórarinsson, F., Magnússon, J. I., & Möller, R. G. (2014). *Úttekt á stærðfræðikennslu í framhaldsskólum [A report on mathematics teaching in upper secondary schools]*. Ministry of Education, Science, and Culture.

Kane, M. T. (2006). Validation. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 17–64). Praeger Publishers.

Kelly, P., Pratt, N., Dorf, H., & Hohmann, U. (2013). Comparing pedagogy in mathematics in Denmark and England. *European educational research journal*, *12*(4), 553–567. https://doi.org/10.2304/eerj.2013.12.4.553

Kilhamn, C., Rystedt, E., Nyman, R., & Holmberg, B. (2018). Video studies in classroom research: Hidden dimensions of teaching and learning. In C. Osbeck, Å. Ingerman, & S. Claesson (Eds.), *Didactic classroom studies: A potential research direction* (pp. 151–171). Nordic Academic Press. https://doi.org/10.21525/kriterium.14.h

Klette, K. (2020). Towards programmatic research when studying classroom teaching and learning. In L. Ligozat, A. Rakhkochkine, & J. Almqvist (Eds.), *Thinking through didactics in a changing world. European perspectives on learning, teaching and the curriculum*. Routledge.

Klette, K., & Blikstad-Balas, M. (2018). Observation manuals as lenses to classroom teaching: Pitfalls and possibilities. *European educational research journal*, *17*(1), 129–146. https://doi.org/10.1177/1474904117703228

Klette, K., Blikstad-Balas, M., & Roe, A. (2017). Linking instruction and student achievement. A research design for a new generation of classroom studies. *Acta Didactica Norge*, *11*(3), 19. https://doi.org/10.5617/adno.4729

Kirsti Klette. (2022). Quality in Nordic teaching (NcoE) common QUINT dataset 2022 (Version 1). Zenodo. https://doi.org/10.5281/zenodo.6381818

Klette, K., Sigurðardóttir, A. K., & Martin, H. (2021). 'Nordic values' and schooling during COVID-19: How to balance comprehensive education and sustainable pandemic regulations. *Education in the North*, *28*(3), 1–4. https://doi.org/10.26203/0855-c687

Klieme, E., Lipowsky, F., Rakoczy, K., & Ratzka, N. (2006). Qualitätsdimensionen und Wirksamkeit von Mathematikunterricht. Theoretische Grundlagen und ausgewählte Ergebnisse des Projekts 'Pythagoras.' In M. Prenzel & L. Allolio-Näcke (Eds.), *Untersuchungen zur Bildungsqualität von Schule. Abschlussbericht des DFG-Schwerpunktprogramms* (pp. 127–146). Waxmann.

Klieme, E., Schümer, G., & Knoll, S. (2001). Mathematikunterricht in der Sekundarstufe I: "Aufgabenkultur" und Unterrichtsgestaltung im internationalen Vergleich [Mathematics instruction at lower secondary level: "Task culture" and quality of instruction in international comparison]. In E. Klieme & J. Baumert (Eds.), *TIMSS – Impulse für Schule und Unterricht: Forschungsbefunde, Reforminitiativen, Praxisberichte und Video-Dokumente* (pp. 43–57). BMBF.

Krauss, S., Bruckmaier, G., Lindl, A., Hilbert, S., Binder, K., Steib, N., & Blum, W. (2020). Competence as a continuum in the COACTIV study: the "cascade model." *ZDM*, *52*(2), 311–327. https://doi.org/10.1007/s11858-020-01151-z

Krummheuer, G. (2002). The comparative analysis in interpretive classroom research in mathematics education. In J. Novotná (Ed.), *European research in mathematics education II: Proceedings of the second conference of the European Society for Research in Mathematics Education*, (pp. 339–346). ERME.

Krummheuer, G. (2007). Argumentation and participation in the primary mathematics classroom. Two episodes and related theoretical abductions. *Journal of mathematical behavior*, *26*(1), 60–82. https://doi.org/10.1016/j.jmathb.2007.02.001

Krummheuer, G. (2012). The "Non-canonical" solution and the "improvisation" as conditions for early years mathematics learning processes: The concept of the "interactional niche in the development of mathematical thinking" (NMT). *Journal für Mathematik-Didaktik*, *33*(2), 317–338. https://doi.org/10.1007/s13138-012-0040-z

Kuhfeld, M. (2017). When students grade their teachers: A validity analysis of the tripod student survey. *Educational assessment*, *22*(4), 253–274. https://doi.org/10.1080/10627197.2017.1381555

Kunter, M., Klusmann, U., Baumert, J., Richter, D., Voss, T., & Hachfeld, A. (2013). Professional competence of teachers: Effects on instructional quality and student development. *Journal of educational psychology*, *105*(3), 805–820. https://doi.org/10.1037/a0032583

Kunter, M., & Voss, T. (2013). The model of instructional quality in COACTIV: A multicriteria analysis. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss, & M. Neubrand (Eds.), *Cognitive activation in the mathematics classroom and professional competence of teachers: Results from the COACTIV project* (pp. 97–124). Springer US.

Kyriakides, L., Christoforou, C., & Charalambous, C. Y. (2013). What matters for student learning outcomes: A meta-analysis of studies exploring factors of effective teaching. *Teaching and teacher education*, *36*, 143–152. https://doi.org/10.1016/j.tate.2013.07.010

Lazarides, R., & Buchholz, J. (2019). Student-perceived teaching quality: How is it related to different achievement emotions in mathematics classrooms? *Learning and instruction*, *61*, 45–59. https://doi.org/10.1016/j.learninstruc.2019.01.001

Leung, D. Y., & Chung, B. P. M. (2019). Content analysis: Using critical realism to extend its utility. In P. Liamputtong (Ed.), *Handbook of research 93methods in health social sciences* (pp. 827–841). Springer. https://doi.org/10.1007/978-981-10-5251-4_102

Liljedahl, P. (2018). Building thinking classrooms: Conditions for problem solving. In P. Felmer, E. Pehkonen, & J. Kilpatrick (Eds.), *Teaching and learning secondary school mathematics* (pp. 307–316). Springer International Publishing. https://doi.org/10.1007/978-3-319-92390-1_29

Liljedahl, P. (2021). *Building thinking classrooms in mathematics, grades K-12: 14 teaching practices for enhancing learning*. Corwin. http://ebookcentral.proquest.com/lib/hi-ebooks/detail.action?docID=6358633

Lipowsky, F., Rakoczy, K., Pauli, C., Drollinger-Vetter, B., Klieme, E., & Reusser, K. (2009). Quality of geometry instruction and its short-term impact on students' understanding of the Pythagorean theorem. *Learning and instruction*, *19*(6), 527–537. https://doi.org/10.1016/j.learninstruc.2008.11.001

Lokan, J., Hackling, M., & Cowan, E. (1999). *Teaching science in Australia: Results from the TIMSS 1999 video study*. Australian Council for Educational Research.

Lopez-Real, F., Mok, I. A. C., Leung, F. K. S., & Marton, F. (2004). Identifying a pattern of teaching: An analysis of a Shanghai teacher's lessons. In L. Fan, N.-Y. Wong, J. Cai, & S. Li (Eds.), *How Chinese learn mathematics: Perspectives from insiders* (pp. 382–412). World Scientific.

Lög um grunnskóla nr. 91/2008 [Law on compulsory schools, no. 91 of 2008]". (2008).

Mahan, K. R., Brevik, L. M., & Ødegaard, M. (2021). Characterizing CLIL teaching: new insights from a lower secondary classroom. *International journal of bilingual education and bilingualism*, *24*(3), 401–418. https://doi.org/10.1080/13670050.2018.1472206

Marton, F., & Booth, S. (1997). *Learning and awareness*. L. Erlbaum Associates.

Mason, J. (2002). *Mathematics teaching practice: A guide for university and college lecturers*. Woodhead.

Mason, J., & Johnston-Wilder, S. (2006). *Designing and using mathematical tasks* (2nd ed.). Tarquin.

McMillan, J., & Schumacher, S. (2014). *Research in education: Evidence-based inquiry*. Pearson.

Mead, M. (1974). Visual anthropology and the discipline of words. In P. Hockings (Ed.), *Principles of visual anthropology* (2nd ed., pp. 3–10). Mouton de Gruyter.

Mok, I. A. C., & Lopez-Real, F. (2006). A tale of two cities: A comparison of six teachers in Hong Kong and Shanghai. In D. Clarke, C. Keitel & Y. Shimizu (Eds.), *Making connections: Comparing mathematics classrooms around the world* (pp. 237–246). Sense.

Morgan, D. (2007). Paradigms lost and pragmatism regained: Methodological implications of combining qualitative and quantitative methods. *Journal of mixed methods research*, *1*(1), 48–76. https://doi.org/10.1177/2345678906292462

Morse, J. M. (2003). Principles of mixed 94methods in94 multimethod research design. In A. Tashakkori & C. Teddlie (Eds.), *Handbook of mixed 94methods in social & behavioral research* (pp. 189–208). Sage.

Neubrand, M., Jordan, A., Krauss, S., Blum, W., & Löwen, K. (2013). Task analysis in COACTIV: Examining the potential for cognitive activation in German mathematics classrooms. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss, & M. Neubrand (Eds.), *Cognitive activation in the mathematics classroom and professional competence of teachers: Results from the COACTIV project* (pp. 125–146). Springer.

Newton, D. P. (2000). *Teaching for understanding*. Routledge.

OECD. (2016). *PISA 2015 results (volume I): Excellence and equity in education*. OECD Publishing. https://doi.org/https://doi.org/10.1787/9789264266490-en

OECD. (2018). What does teaching look like?: A new video study. *Teaching in focus*, *7*(1). https://doi.org/https://doi.org/10.1787/948427dc-en

OECD. (2019). *PISA 2018 results: Combined executive summaries*. https://doi.org/10.1017/CBO9781107415324.004

OECD. (2020). *Global Teaching InSights: A video study of teaching*. OECD. https://doi.org/10.1787/20d6f36b-en

Óskarsdóttir, G. G. (2014). Starfshættir í grunnskólum við upphaf 21. aldar. Háskólaútgáfan.

Österholm, M., Bergqvist, T., Liljekvist, Y., & van Bommel, J. (2021). *The boost for mathematics evaluation report the boost for mathematics evaluation report*.

Otten, S., Webel, C., & de Araujo, Z. (2017). Inspecting the foundations of claims about cognitive demand and student learning: A citation analysis of Stein and Lane (1996). *The journal of mathematical behavior*, *45*(1), 111–120. https://doi.org/10.1016/j.jmathb.2016.12.008

Patton, M. Q. (2015). *Qualitative research & evaluation methods* (4th ed.). Los Angeles, CA: Sage.

Pehkonen, L., Hemmi, K., Krzywacki, H., & Laine, A. (2018). A cross-cultural study of teachers' relation to curriculum materials. In E. Norén, H. Palmér, & A. Cooke (Eds.), *The eighth Nordic conference on mathematics education* (pp. 309–318). Swedish Society for Research in Mathematics Education.

Pehkonen, L., Piht, S., Pakkas, K., Laine, A., & Krzywacki, H. (2018). Estonian and Finnish teachers' views about the textbooks in mathematics teaching. In E. Norén, H. Palmér, & A. Cooke (Eds.), *The eighth Nordic conference on mathematics Education* (pp. 319–328). Swedish Society for Research in Mathematics Education.

Phillips, S. F., Ferguson, R. F., & Rowley, J. F. S. (2021). Do they see what I see? Toward a better understanding of the 7Cs framework of teaching effectiveness. *Educational assessment*, *26*(2), 1–19. https://doi.org/10.1080/10627197.2020.1858784

Pimm, D. (1994). Mathematics classroom language: Form, function and force. In R. Biehler, R. W. Scholz, R. Strässer, & B. Winkelmann (Eds.), *Didactics of mathematics as a scientific discipline* (pp. 159–170). Kluwer.

Pólya, G. (1966). On teaching problem solving. *Conference board of the mathematical sciences: The role of axiomatics and problem solving in mathematics*, 123–129.

Pólya, G. (1981). *Mathematical discovery: On understanding, learning, and teaching problem solving*. Wiley.

Praetorius, A.-K., & Charalambous, C. Y. (2018). Classroom observation frameworks for studying instructional quality: looking back and looking forward. *ZDM*, *50*(3), 535–553. https://doi.org/10.1007/s11858-018-0946-0

Praetorius, A.-K., Klieme, E., Herbert, B., & Pinger, P. (2018). Generic dimensions of teaching quality: the German framework of Three Basic Dimensions. *ZDM*, *50*(3), 407–426. https://doi.org/10.1007/s11858-018-0918-4

Randomness and Integrity Services Ltd. (n.d.). *RANDOM.ORG – True Random Number Service*. Retrieved September 12, 2022, from https://random.org

Rex, L. A., Steadman, S. C., & Graciano, M. K. (2006). Researching the complexity of classroom interaction. In J. L. Green, G. Camilli, & P. B. Elmore (Eds.), *Handbook of complementary methods in educational research* (pp. 727–772). Lawrence Erlbaum.

Rønning, F. (2019). Didactics of mathematics as a research field in Scandinavia. In W. Blum, M. Artigue, M. A. Mariotti, R. Sträßer, & M. V. D. Heuvel-Panhuizen (Eds.), *European traditions in didactics of mathematics* (pp. 153–185). Springer. https://doi.org/10.1007/978-3-030-05514-1_6

Sandilos, L. E., Sims, W. A., Norwalk, K. E., & Reddy, L. A. (2019). Converging on quality: Examining multiple measures of teaching effectiveness. *Journal of school psychology*, *74*(3), 10–28. https://doi.org/10.1016/j.jsp.2019.05.004

Savola, L. (2010). Comparison of the classroom practices of Finnish and Icelandic mathematics teachers. *Journal of mathematics education at teachers college*, *1*(2), 43–55.

Schoenfeld, A. H. (2013). *Classroom observations in theory and practice*. 607–621. https://doi.org/10.1007/s11858-012-0483-1

Schütte, M., Friesen, R., & Jung, J. (2019). Interactional analysis: A method for analysing mathematical learning processes in interactions. In G. Kaiser & N. Presmeg (Eds.), *Compendium for early career researchers in mathematics education* (pp. 101–130). Springer.

Schweig, J. (2014). Cross-level measurement invariance in school and classroom environment surveys: Implications for policy and practice. *Educational evaluation and policy analysis*, *36*(3), 259–280. https://doi.org/10.3102/0162373713509880

Sigurðardóttir, A. K., Guðjónsdóttir, H., & Karlsdóttir, J. (2014). The development of a school for all in Iceland: Equality, threats and political conditions. In U. Blossing, G. Imsen, & L. Moos (Eds.), *The Nordic education model: 'A school for all' encounters neo-liberal policy* (pp. 95–116). Springer.

Sigurjónsson, J. Ö. (2014). *Vitsmunalegar kröfur námsefnis í upprifjunaráföngum framhaldsskóla í stærðfræði: greining á stærðfræðilegum viðfangsefnum í námsefni upprifjunaráfanga*. University of Iceland.

Sigurjónsson, J. Ö., & Kristinsdóttir, J. V. (2018). Upprifjunaráfangar framhaldsskóla í stærðfræði: Skapandi og krefjandi vinna eða stagl? *Icelandic journal of Education*, *27*(1), 65–86. https://doi.org/10.24270/uuom.2018.27.4

Skovsmose, O. (2001). Landscapes of investigation. *ZDM*, *33*(4), 123–132. https://doi.org/10.1007/BF02652747

Smith, E. (2011). Special issue on using secondary data in educational research. *International journal of research & method in education*, *34*(3), 219–221. https://doi.org/10.1080/1743727X.2011.615976

Smith, M. S., & Stein, M. K. (2011). *5 practices for orchestrating productive mathematics discussion*. NCTM.

Statistics Iceland (2020). Grunnskólanemendur eftir bekkjum og skóla 2001-2021 [Number of compulsory school students by grade level and school]. Retrieved September 12, 2022, from https://px.hagstofa.is/pxis/pxweb/is/Samfelag/Samfelag__skolamal__2_grunnskolastig__0_gsNemendur/SKO02102.px.

Stein, M. K., & Smith, M. S. (1998). Mathematical tasks as a framework for reflection: From research to practice. *Mathematics teaching in the middle school*, *3*(4), 268–275.

Stein, M. K., Smith, M. S., Henningsen, M. A., & Silver, E. A. (2009). *Implementing standards-based mathematics instruction: A casebook for professional development* (2nd ed.). National Council of Teachers of Mathematics.

Stigler, J. W., Gonzalez, P., Kawanaka, T., Knoll, S., & Serrano, A. (1999). *The TIMSS videotape classroom study: Methods and findings from an exploratory research project on eighth-grade mathematics instruction in Germany, Japan, and the United States*. National Center for Education Statistics.

Stigler, J. W., & Hiebert, J. (1997). Understanding and improving classroom mathematics instruction: An overview of the TIMSS video study. *Phi Delta Kappan*, *79*(1), 14–21.

Stigler, J. W., & Hiebert, J. (1999). *The teaching gap*. Free Press.

Stovner, R. B., & Klette, K. (2022). Teacher feedback on procedural skills, conceptual understanding, and mathematical practices: A video study in lower secondary mathematics classrooms. *Teaching and Teacher Education*, *110*, 103593. https://doi.org/10.1016/j.tate.2021.103593

Teig, N., Scherer, R., & Nilsen, T. (2019). I know I can, but do I have the time? The role of teachers' self-efficacy and perceived time constraints in implementing cognitive-activation strategies in science. *Frontiers in psychology*, *10*, 1–17. https://doi.org/10.3389/fpsyg.2019.01697

Tekkumru-Kisa, M., Stein, M. K., & Doyle, W. (2020). Theory and research on tasks revisited: Task as a context for students' thinking in the era of ambitious reforms in mathematics and science. *Educational researcher*, *49*(8), 606–617. https://doi.org/10.3102/0013189X20932480

Tengberg, M., van Bommel, J., Nilsberth, M., Walkert, M., & Nissen, A. (2021). The quality of instruction in Swedish lower secondary language arts and mathematics. *Scandinavian journal of educational research*, 1–18. https://doi.org/10.1080/00313831.2021.1910564

Terry, G. (2021). Doing thematic analysis. In E. Lyons, & A. Coyle (Eds.), *Analysing qualitative data in psychology* (pp. 148–161). Sage.

Timmermans, S., & Tavory, I. (2012). Theory construction in qualitative research: from grounded theory to abductive analysis. *Sociological theory*, *30*(3), 167–186. https://doi.org/10.1177/0735275112457914

Ulewicz, M., & Beatty, A. (Eds.) (2001). *The power of video technology in international comparative research in education*. National Research Council. https://doi.org/10.17226/10150

van der Lans, R. M. (2018). On the "association between two things": The case of student surveys and classroom observations of teaching quality. *Educational assessment, evaluation and accountability*, *30*(4), 347–366. https://doi.org/10.1007/s11092-018-9285-5

van der Scheer, E. A., Bijlsma, H. J. E., & Glas, C. A. W. (2019). Validity and reliability of student perceptions of teaching quality in primary education. *School effectiveness and school improvement*, *30*(1), 30–50. https://doi.org/10.1080/09243453.2018.1539015

Voigt, J. (1995). Thematic patterns of interaction and sociomathematical norms. In P. Cobb & H. Bauersfeld (Eds.), *The emergence of mathematical meaning: interaction in classroom Cultures* (pp. 163–202). Lawrence Erlbaum.

von Kotzebue, L., Müller, L., Haslbeck, H., Neuhaus, B. J., & Lankes, E. M. (2020). Cognitive activation in experimental situations in kindergarten and primary school. *International journal of research in education and science*, *6*(2), 284–298. https://doi.org/10.46328/ijres.v6i2.885

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

Wallace, T. L., Kelcey, B., & Ruzek, E. (2016). What can student perception surveys tell us about teaching? Empirically testing the underlying structure of the tripod student perception survey. *American educational research journal*, *53*(6), 1834–1868. https://doi.org/10.3102/0002831216671864

Wang, J., Lin, E., Spalding, E., Klecka, C. L., & Odell, S. J. (2011). Quality teaching and teacher education: a kaleidoscope of notions. *Journal of teacher education*, *62*(4), 331–338. https://doi.org/10.1177/0022487111409551

Weingartner, E. (2021). Cognitive activation potential of E&S tasks at commercial vocational schools in German speaking Switzerland. In M. Blikstad-Balas, K. Klette, & M. Tengberg (Eds.), *Ways of analyzing teaching quality: Potentials and pitfalls* (pp. 204–228). Scandinavian University Press. https://doi.org/10.18261/9788215045054-2021-07

White, M. (2021). A validity framework for the design and analysis of studies using standardized observation systems. In M. Blikstad-Balas, K. Klette, & M. Tengberg (Eds.), *Ways of analyzing teaching quality: Potentials and pitfalls* (pp. 89–120). Scandinavian University Press. https://doi.org/10.18261/9788215045054-2021-03

White, M. C. (2018). Rater performance standards for classroom observation instruments. *Educational researcher*, *47*(8), 492–501. https://doi.org/10.3102/0013189X18785623

Wilhelm, A. G. (2014). Mathematics teachers' enactment of cognitively demanding tasks: Investigating links to teachers' knowledge and conceptions. *Journal for research in mathematics education*, *45*(5), 636–674. https://doi.org/10.5951/jresematheduc.45.5.0636

Willig, G. (2013). *Introducing qualitative research in psychology* (3rd ed.). Open University Press.

Xu, L., & Clarke, D. (2019). Validity and comparability in cross-cultural video studies of classrooms. In L. Xu, G. Aranda, W. Widjaja, & D. Clarke (Eds.), *Video-based research in education: Cross-disciplinary perspectives* (pp. 19–33). Routledge. https://doi.org/10.4324/9781315109213

Yackel, E., & Cobb, P. (1996). Sociomathematical norms, argumentation, and autonomy in mathematics. *Journal for research in mathematics education*, *27*(4), 458–477. https://doi.org/10.2307/749877

Þórðardóttir, Þ., & Hermannsson, U. (2012). *Úttekt á stærðfræðikennslu á unglingastigi grunnskóla*. [A report on mathematics teaching in lower secondary schools]. Ministry of Education, Science, and Culture.

Original Publications

Paper I

Paper II

Paper III

Appendix A – PLATO rubrics

Intellectual Challenge (IC)			
1 **Provides almost no evidence**	**2** **Provides limited evidence**	**3** **Provides evidence with some weaknesses**	**4** **Provides consistent strong evidence**
Teacher provides activities or assignments that are almost entirely rote or recall.	Teacher provides activities or assignments that are largely rote or recall, but a portion of the segment promotes analysis, interpretation, inferencing, or idea generation.	Teacher provides a mix of activity or assignments: most promote analysis, interpretation, inferencing, or idea generation, and a few are focused on recall or rote tasks.	Teacher provides rigorous activities or assignments that largely promote sophisticated or high-level analytic and inferential thinking, including synthesizing and evaluating information and/or justifying or defending their answers or positions.
Adjust up one score point if teacher and student questions and comments are more challenging than the activity as initially presented. High-level questions and comments direct students to: analyze, infer, explain their ideas, or justify their answers. Do not adjust the score if teacher and student questions and comments are in line with rigor of the activity as initially presented to students. Adjust down one score point if teacher and student questions and comments are less challenging than the activity as initially presented. Low level questions and comments direct students to: recall information, restate rote facts, and focus on procedural aspects of a task. Teacher comments that provide "answers" for students also degrade the rigor of the activity.			

Classroom Discourse (CD)				
	1 **Provides almost no evidence**	2 **Provides limited evidence**	3 **Provides evidence with some weaknesses**	4 **Provides consistent strong evidence**
Uptake of student responses	Teacher or students rarely if ever respond to students' ideas about ELA content. Automatic teacher responses that simply acknowledge or echo student contributions (e.g., repetition, "Okay," "Good job," "Thanks") would fall into this category. Teacher accepts answers without asking for clarification or elaboration.	Teacher or students respond briefly to student ideas, and responses do not elaborate or help develop the ideas (e.g., restating without academic language, simple "I agree/disagree" statements that do not specifically reference a previous comment). Alternatively, the teacher may mostly respond to student ideas with automatic responses interspersed with an isolated instance of higher-level uptake (e.g., re-voicing in academic language; asking for clarification, elaboration or evidence).	Teacher or student contributions show a balance between brief responses and higher-level uptake (e.g., re-voicing in academic language; asking for clarification, elaboration or evidence). There are multiple instances in which the teacher or students specifically address student ideas.	Teacher or students consistently engage in high-level uptake of students' ideas, responding in ways that expand on student ideas or enable students to further explain, clarify and specify their thinking.
Opportunities for Student Talk	There are few to no opportunities for mathematics related student talk. Teacher lecture, extended introduction (including giving directions) to an assignment or activity, or recitation formats lasting fewer than 5 minutes would fall in this category.	Talk is tightly teacher-directed, but there are occasional opportunities for brief mathematics related student talk. Examples include recitation formats lasting 5 minutes or longer, or mathematics related talk (whole group, small group, partner talk) lasting fewer than 5 minutes.	Teacher provides opportunities for at least 5 minutes of mathematics related conversation between teacher and students, and/or among students. Some students participate by speaking and/or actively listening, but only 2-3 students are the primary participants. There may still be a substantial amount of teacher direction, and some of the questions that guide the conversation are open-ended. Student-directed discussions that fail to stay on-track would also be at this level.	Teacher provides opportunities for at least 5 minutes of mathematics related conversation between teacher and students, and/or among students. The majority of students participate by speaking and/or actively listening, and students are responding to each other, even if the teacher is still mediating the conversation. The questions that guide the conversation are mostly open-ended, and the focus of the conversation is clear and stays on-track.

Appendix B – The TAG rubric

Task level	Description
Memorisation	- Involve previously learned facts, rules, formulae or definitions being reproduced or committed to memory - Cannot be solved using procedures – either a procedure does not exist or a short time frame does not allow for its use - Are unambiguous in the sense that what is to be reproduced is directly stated and involves exact reproduction of previously seen material - Have no connection to the concepts or meaning that underlie the facts, rules, formulae or definitions being learned or reproduced
Procedures without connections	- Are algorithmic where use of the procedure is either specifically requested or its use is evident based on prior instruction, experience, or placement of the task. - Require limited cognitive demand for successful completion. There is little ambiguity about what needs to be done and how to do it. - Have no connection to the concepts or meaning that underlie the procedure being used - Are focused on producing correct answers rather than developing mathematical understanding - Require no explanations, or explanations that focus solely on describing the procedure that was used
Procedures with connections	- Focus students' attention on the use of procedures for the purpose of developing deeper levels of understanding of mathematical concepts and ideas - Suggest pathways to follow (explicitly or implicitly) that are broad general procedures that have close connections to underlying conceptual ideas as opposed to narrow algorithms that are opaque with respect to underlying concepts - Usually are represented in multiple ways (e.g. visual diagrams, manipulatives, symbols, problem situations). Making connections among multiple representations helps to develop meaning. - Require some degree of cognitive effort. Although general procedures may be followed, they cannot be followed mindlessly. Students need to engage with the conceptual ideas that underlie the procedures in order to successfully complete the task and develop understanding.
Doing mathematics	- Require complex and nonalgorithmic thinking (i.e. there is not a predictable, well-rehearsed approach or pathway explicitly suggested by the task, task instructions, or a worked-out example). - Require students to explore and understand the nature of mathematical concepts, processes, or relationships. - Demand self-monitoring or self-regulation of one's own cognitive processes. - Require students to access relevant knowledge and experiences and make appropriate use of them in working through the task.

Appendix C – Tripod student survey items

Item no.	Item
1)	My teacher in this class makes me feel s/he really cares about me.
2)	My teacher seems to know if something is bothering me.
3)	My teacher really tries to understand how students feel about things.
4)	Student behavior in this class is under control.
5)	I hate the way that students behave in this class. (N)
6)	Student behavior in this class makes the teacher angry. (N)
7)	Student behavior in this class is a problem. (N)
8)	My classmates behave the way my teacher wants them to.
9)	Students in this class treat the teacher with respect.
10)	Our class stays busy and doesn't waste time.
11)	If you don't understand something, my teacher explains it another way.
12)	My teacher knows when the class understands, and when we do not.
13)	When s/he is teaching us, my teacher thinks we understand when we don't. (N)
14)	My teacher has several good ways to explain each topic that we cover in class.
15)	My teacher explains difficult things clearly.
16)	My teacher asks questions to be sure we are following along when s/he is teaching.
17)	My teacher asks students to explain more about the answers they give.
18)	In this class, my teacher accepts nothing less than our full effort.
19)	My teacher doesn't let people give up when the work gets hard.
20)	My teacher wants me to explain my answers—why I think what I think.
21)	In this class, we learn a lot almost every day.
22)	In this class, we learn to correct our mistakes.
23)	This class does not keep my attention—I get bored. (N)
24)	My teacher makes learning enjoyable.
25)	My teacher makes lessons interesting.
26)	I like the way we learn in this class.
27)	My teacher wants us to share our thoughts.
28)	Students get to decide how activities are done in this class.
29)	My teacher gives us time to explain our ideas.
30)	Students speak up and share their ideas about class work.
31)	My teacher respects my ideas and suggestions.
32)	My teacher takes the time to summarise what we learn each day.
33)	My teacher checks to make sure we understand what s/he is teaching us.
34)	We get helpful comments to let us know what we did wrong on assignments.
35)	The comments that I get on my work in this class help me understand how to improve
36)	My teacher moves too fast through the material. (N)
37)	My teacher understands that we may be tired, or that we have had a long day.
38)	My teacher takes time to help each student.

Appendix D – Tripod scales correlations

Table 14. Rank correlations between Reasoning and Discourse scales and the teachers in order by observation evidence.

		T_order	Reasoning	Discourse
T_order	Spearman's rho	–		
	p-value	–		
Reasoning	Spearman's rho	0.093	–	
	p-value	0.186	–	
Discourse	Spearman's rho	-0.027	0.488	–
	p-value	0.705	<0.001	–

Table 15. Intercorrelation (Spearman's rho) between the Reasoning scale and all other items in the Tripod survey.

	Reasoning	All other items
Reasoning	–	–
All other items	0.494*	–

* $p < 0.001$

Table 16. Intercorrelation (Spearman's rho) between the Discourse scale and all other items in the Tripod survey.

	Discourse	All other items
Discourse	–	–
All other items	0.629*	–

* $p < 0.001$

Table 17. Intraclass correlations (ICC) within the Reasoning and Discourse scales.

	ICC1	ICC2
Reasoning	0.112	0.719
Discourse	0.083	0.638

ICC1: Individual-level variance coinciding with group membership

ICC2: Reliability of group means

Appendix E – Participant consent form

QUINT – Quality in Nordic Teaching
Nordic Centre of Excellence

Consent form for students and guardians/parents for participation in the research project QUINT- Quality in Nordic Teaching

We are a group of researchers from the University of Iceland and University of Akureyri. We are a part of group of Nordic research group studying teaching quality across the Nordic classrooms. This project is part of the research conducted by the Nordic Centre of Excellence Quality in Nordic Teaching (QUINT), (see also uio.no/quint), financed by NordForsk and coordinated by the University of Oslo. The main aim of the project is to investigate the different models of classroom instruction in the Nordic classrooms by means of observation and student surveys.

As part of this research project, we would like to ask you for permission to video and audio record lessons in the school subjects Mathematics, Language Arts, and Social Science. We would like to record four consecutive lessons in each subject area this spring. One camera will be placed to record the teacher and the other one will be set up to record students as a class in its entity. The recordings will not disturb the lessons.

This project is abided by and follows current law on processing personal data. We ensure the implementation of necessary procedures compliant with the regulation on data protection and in accordance with the research ethics. Participation in this research is voluntary. There will be no negative consequences if you refuse to participate in the study. The data collected will be kept strictly confidential. We will not include any information in any report we may publish that would make it possible to identify the study participants. Participants' identity will be anonymised in the material that is published.

Sincerely

Anna Kristín Sigurðardóttir University of Iceland
Berglind Gísladóttir University of Iceland
Birna Svanbjörnsdóttir University of Akureyri
Hermína Gunnþórsdóttir University of Akureyri
Kristín Jónsdóttir University of Iceland
Sólveig Zophoníasdóttir University of Akureyri
Rúnar Sigþórsson University of Akureyri
Trausti Þorsteinsson University of Akureyri

HÁSKÓLI ÍSLANDS

Professor Kirsti Klette, QUINT Director

Háskólinn á Akureyri
University of Akureyri

Full name (student)

- ☐ No, I do not wish to take part in the study
- ☐ Yes, I consent to take part in the study descripted above.
  - ☐ Yes, I consent that examples of students' work in the classroom may be collected.
  - ☐ Yes, I consent that parts of the video recordings will be used in teacher training by researchers linked to the Nordic Centre of Excellence (QUINT).
  - ☐ Yes, I consent that parts of the video recordings will be used in professional development of teachers by researchers linked to the Nordic Centre of Excellence (QUINT).
  - ☐ Yes, I consent that parts of the video recordings will be used by researchers linked to the Nordic Centre of Excellence (QUINT).

Date Place

Signature (student)

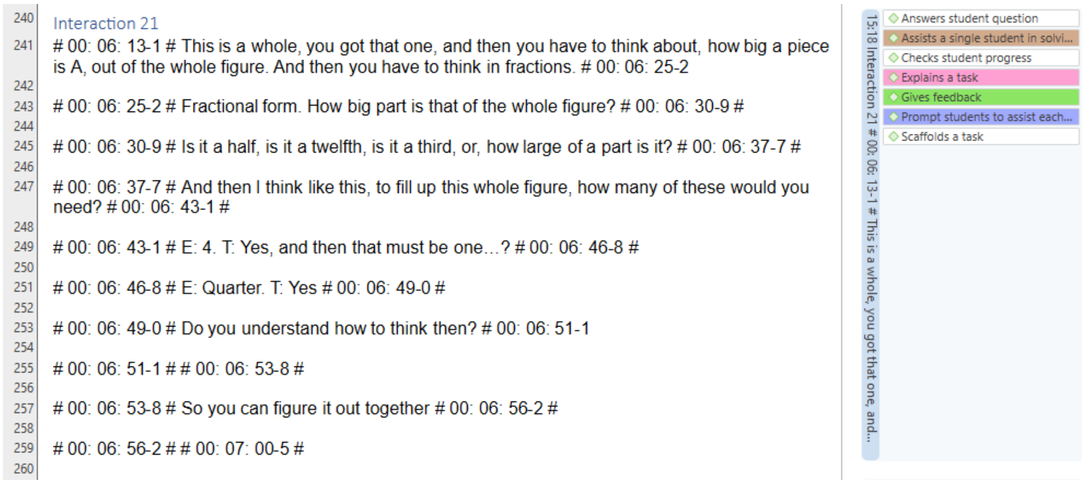
Signature (guardian/parent)

All participation in this research study is voluntary. You may withdraw from the study at any time without affecting your relationship with the investigators of this study. Data obtained in the course of research will be anonymised.

If you have any questions, contact Anna Kristín Sigurðardóttir í HÍ, aks@hi.is, P: 692-2742

Please submit the signed form to your teacher as soon as possible.

Appendix F – Data examples of coded interactions



Interaction 21

00: 06: 13-1 # This is a whole, you got that one, and then you have to think about, how big a piece is A, out of the whole figure. And then you have to think in fractions. # 00: 06: 25-2

00: 06: 25-2 # Fractional form. How big part is that of the whole figure? # 00: 06: 30-9

00: 06: 30-9 # Is it a half, is it a twelfth, is it a third, or, how large of a part is it? # 00: 06: 37-7

00: 06: 37-7 # And then I think like this, to fill up this whole figure, how many of these would you need? # 00: 06: 43-1

00: 06: 43-1 # E: 4. T: Yes, and then that must be one…? # 00: 06: 46-8

00: 06: 46-8 # E: Quarter. T: Yes # 00: 06: 49-0

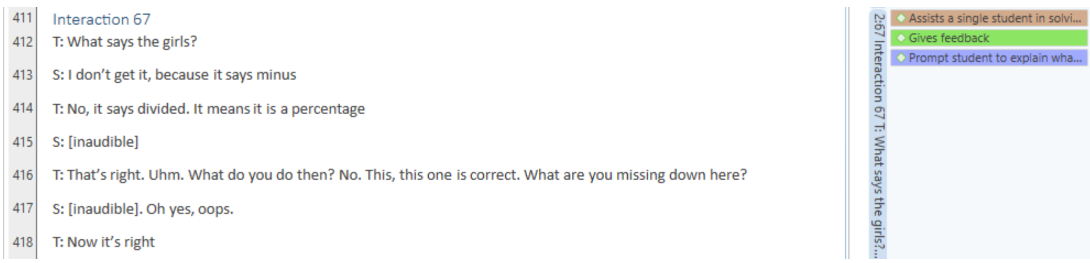
00: 06: 49-0 # Do you understand how to think then? # 00: 06: 51-1

00: 06: 51-1 # # 00: 06: 53-8

00: 06: 53-8 # So you can figure it out together # 00: 06: 56-2

00: 06: 56-2 # # 00: 07: 00-5

Figure 12. Data example of a coded interaction in a lesson from Sweden.



Interaction 67

T: What says the girls?

S: I don't get it, because it says minus

T: No, it says divided. It means it is a percentage

S: [inaudible]

T: That's right. Uhm. What do you do then? No. This, this one is correct. What are you missing down here?

S: [inaudible]. Oh yes, oops.

T: Now it's right

Figure 13. Data example of a coded interaction in a lesson from Denmark.

Appendix G – Development of themes

Initial candidate theme	Frequent shifts between prompting, feedback, and explanation	Connections within mathematics and to non-mathematical experience	"Withitness" as both professional flexibility and caring
Associated codes	Prompts student to explain "how" Prompts student to explain what they are doing Gives feedback Gives a hint Explains a game/group activity Explains a method	Connects to past experience/activity Connects math to daily life Connects one method to another Connects one concept to another	Checks student progress Checks student feelings Supports student agency Withitness

Final theme	Frequent shifts in types of interactions	Connections within mathematics and to non-mathematical experience	Use of formative feedback and explicit student roles
Associated codes	Prompts student to explain "how" Prompts student to explain what they are doing Gives feedback Gives a hint Explains a game/group activity Explains a method	Connects to past experience/activity Connects math to daily life Connects one method to another Connects one concept to another	Gives feedback Formative feedback Assigns a role to student(s) Checks student progress