Detect, diagnose and intervene

STUDENTS' MATHEMATICS-SPECIFIC DIFFICULTIES WITH LINEAR EQUATIONS IN AN ONLINE LEARNING ENVIRONMENT

Morten Elkjær – PhD Dissertation
AARHUS UNIVERSITY

PhD student: Morten Elkjær

Principal supervisor: Professor Uffe Thomas Jankvist, Aarhus University, Denmark

Co-supervisor: Professor Jeremy Hodgen, University College London, England

Company supervisor: Klaus Bruun Pedersen, ALINEA, Denmark

ISBN: 978-87-7507-563-8

DOI: 10.7146/aul.530

This PhD dissertation consists of the following six paper contributions and a linking text (a *kappa*) that unifies the six papers via a binding methodology.

Paper A **Elkjær, M., & Hansen, C.** (2020). Clustering student errors. In B. Barzel, R. Bebernik, L. Göbel, M. Pohl, H. Ruchniewicz, F. Schacht, & D. Thurm (Eds.), *Proceedings of the 14th International Conference on Technology in Mathematics Teaching – ICTMT 14, Essen, Germany,* 165–172.

Paper B **Elkjær, M., & Jankvist, U. T.** (2021). Designing tasks for a dynamic online environment: Applying research into students' difficulties with linear equations. *Mathematics*, *9*(5), 557. https://doi.org/10.3390/math9050557

Paper C **Elkjær, M., & Hodgen, J.** (2022). Operationalising Vergnaud's notion of scheme in task design in online learning environments to support the implementation of formative assessment. *Implementation and Replication Studies in Mathematics Education*, *2*(1). https://doi.org/10.1163/26670127-bja10002

Paper D **Elkjær, M., & Hodgen, J.** (2022). Replicating a study with tasks associated with the equals sign in an online environment. In C. Fernández, S. Llinares, A. Gutiérrez, & N. Planas (Eds.), *Proceedings of the 45th Conference of the International Group for the Psychology of Mathematics Education*: *Vol. 2* (pp. 251–258). PME.

Paper E **Elkjær, M., & Mørup, M.** (submitted). Learning from errors: Co-clustering students' answer types in a digital learning environment to verify task design on linear equations.

Paper F **Elkjær, M., & Thomsen, L. A.** (2022). Adapting the balance model for equation solving to virtual reality. *Digital Experiences in Mathematics Education*. Springer https://doi.org/10.1007/s40751-022-00103-4

Acknowledgements

I wish to express my sincere gratitude to all who have made this PhD project possible. I am deeply grateful to the Innovation Fund Denmark and Edulab (ALINEA), who financially have made everything possible.

I have been lucky to be a part of what has felt like a team of PhD students in the department of Educational Theory and Curriculum Studies at the Danish School of Education. I have been fortunate enough to have worked with inspiring, versatile and clever fellow PhD students in and around our 'Thurday meetings'. Thank you all for the discussions and times we have shared.

I want to thank the team at MatematikFessor, both from Edulab and ALINEA. Together you have provided me with a unique opportunity to investigate a rare and underutilised source of educational data. I would like to express my gratitude to Kasper Holst Hansen for believing in the project and for being willing to bridge the gap between research and the industry of developing teaching materials. I am deeply grateful to my company supervisor Klaus Bruun Pedersen, for always being on my side. You have provided me with assistance in every aspect of accessing tasks and the acquiring the data. My queries or problems never felt unimportant and I have always trusted that you would and could help my ideas come to life.

I would like to express my sincere gratitude to my co-supervisor Prof. Jeremy Hodgen and Prof. Dietmar Küchemann. You gave me a valuable opportunity to study in another research environment. You also provided me with unique insight to the field of diagnostic task design in mathematics and an extended insight into interpreting students' errors. Jeremy you gave me a lot of your time and I truly appreciate our discussions and our writing processes.

Tack Ola Helenius and Linda Ahl for instructing me to the Theory of Conceptual Fields. This way of describing knowledge, actions and connections between subject, teacher and learner has been of great value to me.

I am deeply grateful to Prof. Morten Mørup. Thank you for instantly agreeing to apply you time and field of expertise to analyse my data. You provided me with a chance to change research environment and knowledge about what machine learning can do for educational data.

Thank you Uffe – for everything. None of this would have been possible without you. You are truly an expert – an expert researcher and an expert supervisor. For the countless questions I have presented you with and ideas I have unloaded on you. But mostly, for letting me do what I wanted and for making coffee while playing cards.

Finally, to my family, thank you. I love you more than you could ever know. With all my love, thank you Mai, Villy and Ditlev. One and a half years delay, but we finally arrived. I love you and thank you for always reminding me of what really matters.

Resumé på dansk

Denne ph.d.-afhandling udforsker de matematik-specifikke vanskeligheder, studerende står over for, når de arbejder med lineære ligninger i et online læringsmiljø kaldet MatematikFessor. Miljøet har eksisteret i mere end 12 år og giver elever mulighed for at arbejde med matematiske opgaver i en webbrowser. Lærere kan tildele deres elever arbejde i miljøet, og miljøet præsenterer lærerne for statistiske overblik baseret på, hvor mange rigtige svar eleverne afgiver.

Mit formål med at udforske data fra MatematikFessor er at give lærere nye muligheder for at hjælpe deres elever med at overvinde de matematiske specifikke vanskeligheder der opleves i arbejdet med lineære ligninger. Rammerne for at arbejde med elever, der oplever vanskeligheder, udspringer af Matematikvejlederuddannelsens ramme (Jankvist & Niss, 2015) for at detektere elever, der oplever vanskeligheder og diagnosticere oprindelsen eller årsagerne til disse vanskeligheder, inden der etableres interventioner, der skal hjælpe eleverne med at komme forbi disse vanskeligheder.

Afhandlingen består af en samling af seks artikler og dette dokument, der fungerer som en metodisk struktur for det samlede ph.d.-projekt.

Artikel A er skrevet i samarbejde med med-ph.d.-studerende Christian Hansen og havde til formål at verificere det diagnostiske potentiale af opgaver fokuseret på ligningsløsning, der allerede var implementeret i MatematikFessor. Vi fandt ud af, at disse opgaver ikke er tilstrækkelige til at muliggøre identifikation af elevernes vanskeligheder med at fortolke og løse lineære ligninger.

Artikel B fungerer som rygraden i denne afhandling og er skrevet i samarbejde med min hovedvejleder prof. Uffe Thomas Jankvist. Denne artikel præsenterer en forskningslitteraturgennemgang om elevers vanskeligheder med at arbejde med begreber tilknyttede lineære ligninger. Endvidere præsenterer artiklen et sæt designprincipper baseret på litteraturgennemgangen til at designe opgaver, der sigter mod at afdække elevernes vanskeligheder med at løse lineære ligninger, egnet til implementering i MatematikFessor.

Artikel C er skrevet i samarbejde med min medvejleder Prof. Jeremy Hodgen og præsenterer designprincipper for design af alternative opgaver til online læringsmiljøer, der sigter mod at sætte lærere i stand til at stille hypoteser om deres elevers skemaer (Vergnaud, 2009) som et udtryk for deres handlinger når der arbejdes med elementer af ligningsløsning, mere præcist fortolkningerne af lighedstegnet.

Artikel D er også skrevet i samarbejde med med-vejleder prof. Jeremy Hodgen og giver indsigt i, hvor vigtig opgavedesign er, når opgaver designes til diagnostiske formål. Artiklen præsenterer et replikationsstudie med den berømte opgave $8 + 4 = \_\_ + 7$ (Falkner et al., 1999), inklusiv variationer implementeret i MatematikFessor. Resultaterne fra den oprindelige undersøgelse blev sammenlignet med resultaterne fra Paper D og giver anledning til interessante forskelle og ligheder.

Artikel E er skrevet i samarbejde med professor Morten Mørup (Danmarks Tekniske Universitet) og præsenterer en storstilet dataanalyse af 892 unikke ligninger designet efter principperne i Paper B. Analysen benyttede et omfattende kodesystem baseret på fortolkningen af de fem mest populære svar som svar på hver af de 892 ligninger. Undersøgelsen brugte co-clustering til at observere grupperinger af både elever og opgaver, der udviser lignende adfærd. Dataene består af 2.135.968 unikke svar leveret af 94.368 elever i MatematikFessor. Artiklen giver indsigt, der muliggør etablering af midler til at diagnosticere elever, der har problemer med at arbejde i matematikfessor, ved at generere omfattende feedback til lærere om deres elevers fejl og årsagerne til dem. Artiklen afslørede grupper af opgaver, der er særligt velegnede til at afsløre en bestemt fejltype.

Artikel F er skrevet i samarbejde med med-ph.d.-studerende Lui A. Thomsen og er et forsøg på at designe og udvikle et virtuelt miljø, der giver lærere og studerende mulighed for at arbejde med at løse ligninger sammen.

Miljøets specifikationer gør det muligt for brugere at arbejde med negative tal på den klassiske balancemodel, da tyngdekraften, der påføres af negative mængder på vægten, kan arbejde efter hensigten til at trække op på balanceskålen i stedet for at skubbe ned, som normale vægte ville.

Overordnet præsenterer afhandlingen viden om, hvad litteraturen og udforskningen af data fra MatematikFessor kan afsløre om de vanskeligheder, eleverne møder, når de arbejder med lineære ligninger. Indførelsen af terminologien for en opgaves 'diagnostiske værdi' bidrager til ideen om at undersøge kvaliteten af opgavedesign til diagnostiske formål. Derudover præsenterer afhandlingen forskningsbaserede forslag til, hvordan lærere kan lære af deres elevers fejl og deres vanskeligheder med at løse ligninger via forbedret feedback fra online læringsmiljøer. Denne afhandling foreslår, at udviklere af online læringsmiljøer deler ansvaret for at fortolke elevernes svar for at gøre det muligt for lærere bedre at hjælpe deres elever med at lære om lineære ligninger. Endvidere præsenterer afhandlingen ideer til, hvordan lærere kan arbejde sammen med deres elever om at overvinde elevens matematikspecifikke vanskeligheder relateret til negative tal, når der løses ligninger i et virtuelt miljø.

Didactical Engineering fungerer som en metodisk struktur for dette denne del af afhandlingen, der præsenterer og diskuterer resultaterne i de seks inkluderede artikler. Opbygningen af *Didactical Engineering* fører læseren gennem dokumentet i fire overordnede faser. Efter introduktionen og præsentationen af forskningsproblematikken er ph.d.-projekt skriftlige bidrag struktureret i de fire faser af *Didactical Engineering*.

Summary in English

This PhD dissertation explores the mathematics-specific difficulties students face when working with linear equations in an online learning environment called MatematikFessor. The environment has existed for more than 12 years and allows students to work on mathematical tasks in a web browser. Teachers can assign their students work in the environment, and the environment in turn presents teachers with statistical overviews based on how many correct answers students provide.

My purpose for exploring the data from MatematikFessor was to provide teachers with new opportunities to help their students overcome their mathematic specific difficulties working with linear equations. The framework for working with students experiencing difficulties stems from the Maths Counsellor Programme's framework (Jankvist & Niss, 2015) for detecting students experiencing difficulties and diagnosing the origin or causes of these difficulties before establishing interventions to help students move past these difficulties.

The dissertation consists of a collection of six papers and this document, which serves as a methodological structure for the overall PhD project.

Paper A was written in collaboration with fellow PhD student Christian Hansen and set out to verify the diagnostic potential of tasks focused on solving linear equations already implemented in matematikfessor. We found that these tasks are not sufficient in enabling the identification of students' difficulties in interpreting and solving linear equations.

Paper B serves as the backbone of this dissertation and was written in collaboration with my principal supervisor Prof. Uffe Thomas Jankvist. This paper presents a literature review on students' difficulties working with the concept of linear equations. Furthermore, the paper presents a set of design principles based on the literature review for designing tasks aimed at exposing students' difficulties in solving linear equations suited for implementation in MatematikFessor.

Paper C was written in collaboration with my co-supervisor Prof. Jeremy Hodgen and presents design principles for designing alternative tasks for online learning environments aimed at enabling teachers to hypothesise about their learners' schemes (Vergnaud, 2009) as an expression of their actions working with elements of equation solving, more precisely the interpretations of the equals sign.

Paper D was also written in collaboration with co-supervisor Prof. Jeremy Hodgen and provides insight into how important task design is when designing tasks for diagnostic purposes. The paper is a replication study of the famous task $8 + 4 = \_\_ + 7$ (Falkner et al., 1999), with variations implemented in matematikfessor. Findings from the original study were compared with those of Paper D and presents important differences and similarities.

Paper E was written in collaboration with Professor Morten Mørup (Technical University of Denmark) and presents a large-scale data analysis of 892 unique equations designed using the principles in Paper B. The analysis utilised an extensive coding system based on the interpretation of the five most popular answers in response to each of the 892 equations. The study used co-clustering to observe groupings of both students and tasks displaying similar behaviours. The data consist of 2,135,968 unique answers provided by 94,368 students in matematikfessor. The paper provides insight enabling the establishment of means for diagnosing students facing difficulty working in matematikfessor by generating extensive feedback for teachers about their learners' errors and reasons for them. The paper unveiled groups of tasks that are particularly suited to revealing a particular error type.

Paper F was written in collaboration with fellow PhD student Lui A. Thomsen and is a novel attempt at designing and developing a virtual environment that allows teachers and students to work on solving equations together. The specifics of the environment allow users to work with negative quantities on the classical balance

model, since the gravitational force applied by negative quantities to the balance can work as intended to pull up on the balance pan instead pushing down as normal weights would.

Overall, the dissertation presents knowledge on what the literature and exploration of data from matematikfessor can reveal about the difficulties students face when working with linear equations. The introduction of the terminology of the 'diagnostic value' of a task contributes to the idea of investigating the quality of task design for diagnostic purposes. Additionally, the dissertation presents research-based suggestions on how teachers can learn from their students' errors and their difficulties solving equations via improved feedback from online learning environments. This dissertation suggests that developer of online learning environments share responsibility in interpreting students' answers in order to enable teachers to better help their students learning about linear equations. Furthermore, the dissertation presents ideas on how teachers can work together with their students to overcome mathematics specific difficulties related to negative numbers when solving equations in a virtual environment.

Didactical engineering serves as a methodological structure for this document presenting and discussing the findings in the six paper contributions included. The structure of didactical engineering leads the reader through the document in four overall phases. After the introduction and the presentation of the research *problématique*, the research is structured in the four phases of didactical engineering.

Table of Contents

List of figures and tables

Chapter 1 Introduction and research *problématique*

Students experience difficulties when introduced to the components of algebra in the transition to generalised arithmetic (Rhine et al., 2018). In particular, the introduction of the concept of linear equations and the shift in the meanings of accompanying concepts, such as the equals sign and presence of letters as mathematical objects, create confusion (Kieran, 1981; Küchemann, 1981; MacGregor & Stacey, 1993; Matthews et al., 2012; Rhine et al., 2018; Vlassis, 2002).

In the recent decade, the transition to the inclusion of digital environments and resources in the classroom has been swift and continues to remain a prevalent phenomenon. Such learning environments (digital textbooks and/or resources) promise to better enable teachers to monitor their students' performance, to improve their teaching and ultimately to enhance student learning. Mathematics education was one of the first fields to recognise the potential of digital technologies and explore how these can improve teaching and learning, as well as how these could be embedded in mathematics curricula (Trouche et al., 2013). Digital learning environments and digital resources present a number of potential data-driven benefits (Dawson et al., 2018; Spitzer & Musslick, 2021). The content in the environments can be tailored to the learning needs of students by their teachers, by students themselves or 'intelligently' in that the environment can dynamically provide tasks based on students' previous responses, introducing what is known as individualised or personalised learning (Steenbergen-Hu & Cooper, 2013). Moreover, these learning environments can score students' responses almost immediately, thus providing automated feedback for students and teachers (Cavalcanti et al., 2021; Dawson et al., 2018; Rezat, 2021).

In recent years, we in the field mathematics education research have become familiarised with the fact that teachers struggle to engage with or make use of the formative values or possibilities generated by digital learning environments and digital resources in terms of the feedback teachers get on their students (Utterberg Modén, 2021). Additionally, teachers find that the dashboards of various digital learning environments only provide information about their students' performance and not their understanding of mathematical concepts and general mathematical abilities, and creating teaching plans with a digital mathematics textbooks was found to increase teachers' workloads (Utterberg Modén, 2021). Furthermore, teachers need additional competencies and time for teaching when initially implementing such resources. Teaching based on data from digital resources and adaptive functions, such as adaptive task selection for students, builds on a personalised learning approach. Utterberg Modén (2021) emphasised that this contradicts established teaching norms of building on collective classroom activities, situations where most activities are organised by the teacher to develop not only mathematics knowledge and abilities but also more general skills.

In a recent review of teachers' use of student data in the classroom, Sun et al. (2016) concluded that even though teachers generally have a positive attitude towards the functionalities enabled by data, they often expressed that a lack of time, difficulties with data and data tools, relevance and usefulness and other factors hindered their engagement in data use. Furthermore, a review of the use of formative feedback found that feedback lacking specificity can cause confusion instead of offering the intended help to teachers (Shute, 2008).

Prior studies have revealed that modern teachers, utilising the affordances enabled by digital resources and learning environments, do not find the automated feedback they are presented with appropriate. Thus, the question becomes how should developers of digital learning environments and other digital resources for mathematics learning engage in the development of this feedback to exploit the potential within these powerful online learning environments? Furthermore, the appropriateness of feedback could also be related to the *domain dependency issue* raised in a critical review of formative assessment and its use in the classroom (Bennett, 2011):

> To be maximally effective, formative assessment requires the interaction of general principles, strategies, and techniques with reasonably deep cognitive-domain understanding. That deep cognitive-domain understanding includes the processes, strategies and knowledge important for proficiency in a domain, the habits of mind that characterise the community of practice in that domain, and the features of tasks that engage those elements. (Bennett, 2011, p. 15)

In fact, here is where this discussion regarding the two approaches of offering feedback to teachers and the idea of formative assessment coincide. When discussing automated feedback provided by digital environments or resources on the mathematics-specific learning difficulties resulting in student errors, the feedback might not have been instigated by the teachers themselves in the same way formative assessment might initially be planned and performed. Could one assume that teachers initiating work in digital learning environments or digital resources expect the automated assessment of the students' work? Wiliam and Thompson (2007) advocated for a 'big idea' in formative assessment: 'The "big idea" is that evidence about student learning is used to adjust instruction to better meet student needs—in other words that teaching is adaptive to the student's learning needs' (p. 15). Wiliam and Thompson (2007) further summarised formative assessment as comprising the following five key strategies (p. 15):

1. Clarifying, understanding and sharing learning intentions
2. Engineering diagnostic tasks that elicit evidence of learning
3. Providing feedback that moves teaching forward
4. Activating students as learning resources for one another
5. Activating students as owners of their own learning

I realise that there is an important part of formative assessment that revolves around how evidence about student learning is transformed into feedback for students. However, it is equally important for teachers to access this evidence. Additionally, Bennet (2011) advocated for several other issues, including the *measurement issue*; he argues that assessment is not simply the process of observing students' responses and noting errors or difficulties. Rather, it is an inferential process that requires teachers or researchers to have substantial knowledge and expertise that enables them to make productive 'formative hypotheses' and then to act on these. Bennett (2011) argued that this may involve engaging with a student to probe why they gave a particular answer. Additionally, the teacher could assign more tasks in an attempt to determine a pattern in the answers consistent with the hypothesis. This idea leads well to the introduction of the *Maths Counsellor Programme* developed in Denmark by Prof. Jankvist and Prof. Niss (Jankvist & Niss, 2015). This programme is not necessarily a revolutionary way of working to overcome students' difficulties in mathematics education; however, they proposed a sensible framework for addressing and counteracting students' mathematics-specific difficulties (Jankvist & Niss, 2015). The framework, if you will, uses diagnostic items to detect students experiencing difficulties before establishing a diagnostic session (Jankvist & Niss, 2017). Based on the diagnostic process, an intervention is designed to counteract some of the more or less rigid understanding causing the concrete difficulties.

Providing high-quality feedback about students' understanding or general abilities through digital learning environments does not seem at all straightforward. However, as digital learning environments become more ubiquitous, it is crucial to develop an explicit understanding of the process of designing tasks for implementation in digital learning environments to enable and support teachers in applying the 'big idea' in preparing better teaching (Wiliam & Thompson, 2007).

1.1 Personal motivation building up to the *problématique*

While writing my master's thesis on mathematics education, I worked towards achieving quantitative insight into the mathematics-specific difficulties experienced by upper secondary school students working with linear equations. In 2017, I was asked if I wanted to pursue the academic experience of writing a PhD dissertation with similar opportunities to investigate data. The PhD project was proposed to me as an industrial PhD project, where I would work with my supervisor Prof. Uffe Thomas Jankvist and Edulab, the company that built and maintains an online learning environment for mathematics accessible through a webpage (www.matematikfessor.dk). I had worked with Prof. Jankvist during and after my master's thesis, and I was familiar with this online learning environment that Edulab had developed.

During my master's studies, I worked with data from the Maths Counsellor Programme that was developed by Prof. Niss and Prof. Jankvist (2015). This programme was designed to give upper secondary mathematics teachers insight into how the research field of math education could benefit their work by helping students at their respective educational institutions overcome mathematics-specific difficulties. The programme would result in teachers being educated as maths counsellors. Later, I return to the reasoning for why I am primarily using the term 'difficulties' when discussing students' problematic experiences with or problematic conceptions of the elements of mathematics. The programme's framework for working with students with difficulties in mathematics is broken down into three overarching parts: *detect*, *diagnose* and *intervene* (Jankvist & Niss, 2015). The focus of the teachers in the programme is to detect students experiencing difficulties and then diagnose the origin or the cause of these particular difficulties before carrying out interventions to hopefully help students progress in their learning and move on from the particular difficulties.

My work with the data from the Maths Counsellor Programme led to some interesting findings because of the shift from the intended qualitative perspective to a quantitative perspective. Prof. Jankvist and Prof. Niss, who are authors of the so-called 'detection test' that are utilised in the Maths Counsellor Programme, were interested in the added insight into what the tasks were capable of illuminating based on the quantitative analyses of the data. Therefore, my master's project inspired my interest in task design with a detection or a diagnostic (or formative) purpose. Several years after finishing my master's, I was asked if I wanted to apply for an industrial PhD position offering the opportunity to generate and analyse large amounts of data, thus reigniting my interest. The knowledge gathered from my master's project confirmed what we might have already known or suspected—that upper secondary school students in Denmark experience learning difficulties related to algebraic expressions and equations.

I have always had a passion for working with or teaching the concept of linear equations. After realising through my master's project that so many Danish students struggle with learning the concept and that these struggles persist into upper secondary school, I wanted to explore ways I could help students learn the concept through data investigation. I was excited to work with lower secondary school students as one of the stakeholders, since their difficulties learning the concept could be addressed, possibly smoothening the transition from lower secondary to upper secondary school.

1.2 Employer's motivation building up to the *problématique*

For many years, Edulab had been the lead distributor of online learning environments for mathematics in Denmark, despite being a small independent company. Within the first one and a half years of my employment as an industrial PhD student, the company faced a situation that led to major organisational rearrangements, and the staff was approximately reduced by half. This, among other things, led to the company being sold to a large publisher of both online and analogue teaching materials for Danish schools, called ALINEA. Edulab is still the provider of MatematikFessor (slang for mathematics professor), the largest online learning environment for mathematics in Denmark.

Through my employment with Edulab, I became familiar with unique ways to extract data and the generous opportunities for generating new means to collect data that I could investigate to learn about the students and teachers using the online learning environment. The potential was there; I just needed to establish myself as a PhD student to learn what kind of data I would actually need and how I could manage this learning process in a fruitful way. Edulab was excited to have a research-focused person explore the possibilities of learning from or utilising the data they produce. At the time of my employment, Edulab was already managing and further developing an adaptive algorithm to help students engage with content that fits their level of performance and adapts to their skills in mathematics. The idea that I would be able to establish additional insights to refine this algorithm was also among the initial interests of the company. Additionally, they were interested in establishing a connection to the math education research field and to develop their content based on findings and trends from the field.

1.3 Research problématique

The aim of the present industrial PhD research project naturally became a constellation based on the motivations of the different stakeholders in the project. As a host and co-founder of the project, Edulab gave me free reign on how to establish a sound research project based on the possibilities and data that they were able to provide through their online learning environment. The plan from the beginning was that the mathematical topic should be centred on the concept of linear equations. However, instead of upper secondary school students, like those observed in my master's research, this PhD project focused on lower secondary school students and linear equations appropriate to the lower secondary level in Denmark. This sounded very interesting to me, since I had learned that many of the difficulties faced by upper secondary school students originated in primary and lower secondary school.

The Maths Counsellor Programme inspired my approach when conceptualising this PhD project and framing the ideas of my investigation into students' learning in online learning environments. In the Maths Counsellor Programme, teachers are inspired to follow the established framework for addressing students' difficulties in their mathematics classroom. Specifically, they are asked to engage in the *detection* of students experiencing difficulties or obstacles in learning mathematics, in the *diagnosis* of the origin or cause of these observed difficulties and finally in designing an *intervention* directed towards the diagnosed difficulties (Jankvist & Niss, 2015). I note here that this procedure of detection, diagnosis and intervention is mainly possible when personal and face-to-face interaction is possible. Via students' engagement with or in online learning environments, developers, teachers and researchers do not have access to this same interaction.

Earlier, I established that research into the feedback offered by online learning environments and digital resources to teachers is insufficient. Because of the way online learning environments are typically structured, with limited input types in many cases restricted to an input field and multiple-choice answers, I wanted to explore the potential for improved feedback for teachers using online learning environments in their classrooms. When I say improved feedback, I specifically mean that online learning environments traditionally do not provide teachers with detailed descriptions, assumptions or means to hypothesise about their students' mathematics-specific difficulties; rather, they provide statistical insights into how the score distributions of the class are based on the tasks students have completed. I return to this during the institutional analysis (section 2.1).

Additionally, I wanted to find evidence to support how important task design is when the task is supposed to generate data for unveiling students' difficulties related to solving linear equations and ultimately provide feedback about these difficulties that is more effective than a performance measure. I imagined that because the input types offered by online learning environments are restricted, a heavy focus on high-quality task design should be the focus of designers.

Finally, I wanted to explore how an intervention could be designed to counteract students' known difficulties with the concept of linear equations. I wanted to create/design an opportunity for teachers to join their students in a guided intervention. When I joined Edulab as an industrial PhD student, I was introduced to two fellow PhD students already employed by Edulab. They were Christian Hansen (computer science) who was working with machine learning algorithms and Lui A. Thomsen (human–computer interaction) who was working with learning utilising virtual reality (VR). It became a goal for me to explore opportunities for collaboration with my then newfound colleagues. I expected that if I wanted to explore the possibilities of data-driven feedback generated for teachers, collaboration with skilled computer scientists might be necessary. Additionally, I wanted to establish grounds for different fields of research to merge in an attempt to create better or more interesting results regarding interventions to help students improve their linear equation and equation-solving abilities.

In the following I present the research questions I formulated in the endeavour to explore the *problématique* presented here. Hereafter, I present the reader with the methodological framework that serves as the structure for this *kappa* (I use the Swedish word 'kappa' for linking text) constitute my PhD dissertation.

1.4 Research questions

Based on the introduced motivations for the project valued by the respective stakeholders and the established *problématique*, I formulated the research aim into specific research questions. In posing these research questions, I remind the reader that they were heavily inspired by the framework for working with mathematics-specific difficulties established by the Maths Counsellor Programme (Jankvist & Niss, 2015). The procedure, as mentioned, involves detection, diagnosis and intervention and should not necessarily be converted one to one to fit the situation of working with online learning environments. Rather, I hope for the reader to become familiar with the origin of the procedure and the inspiration for the present study's framework.

1. *What design principles are appropriate when structuring and designing tasks capable of detecting students facing mathematics-specific difficulties related to learning the concept of linear equations and equation solving?*
    a. *What principles are appropriate when exploring one specific difficulty (the equals sign)?*
    b. *What principles are appropriate when exploring difficulties related to the more general topic of equation solving found in the literature?*
2. *What possibilities for the general diagnosis of students' difficulties related to equation solving can be established?*
    a. *What possibilities can be established based on a task found in the relevant literature?*
    b. *What possibilities can be established based on the analysis of an exhaustive set of tasks involving equation solving?*
3. *What possibilities for general interventions in relation to the concept of linear equations and equation solving can be established?*
    a. *What possibilities can be established based on a specific difficulty from the literature?*
    b. *What possibilities can be established based on difficulties measured in an exhaustive set of tasks involving equation solving?*

The proposed research questions are explored retrospectively through this *kappa* and qualified through the included papers using DE. A commonality linking all the included papers is that digital task design as a qualifier for teaching design plays a significant role. In the following, I go through the phases in the DE process and establish a place for and the relevance of the included papers in each phase.

1.5 Methodological considerations

In the following sections, I establish the methodological foundation for the present research project. The project mainly sought to design various initiatives to answer the research questions that emerged from the *problématique*. Didactical engineering (DE) (Artigue, 2009, 2014, 2015) became the structural research methodology I ultimately chose. DE emerged from French didactical culture of the early 1980s. DE has been developed in close relation to the theory of didactical situations (TDS) (Brousseau, 2006).

DE is structured into four different phases: *preliminary analyses*; *conception and a priori analysis*; *realisation, observation and data collection*; and *a posteriori analysis and validation* (Artigue, 2015). The respective phases include specific prescribed dimensions that demand attention. An overview of DE as a research method is depicted in Figure 1.

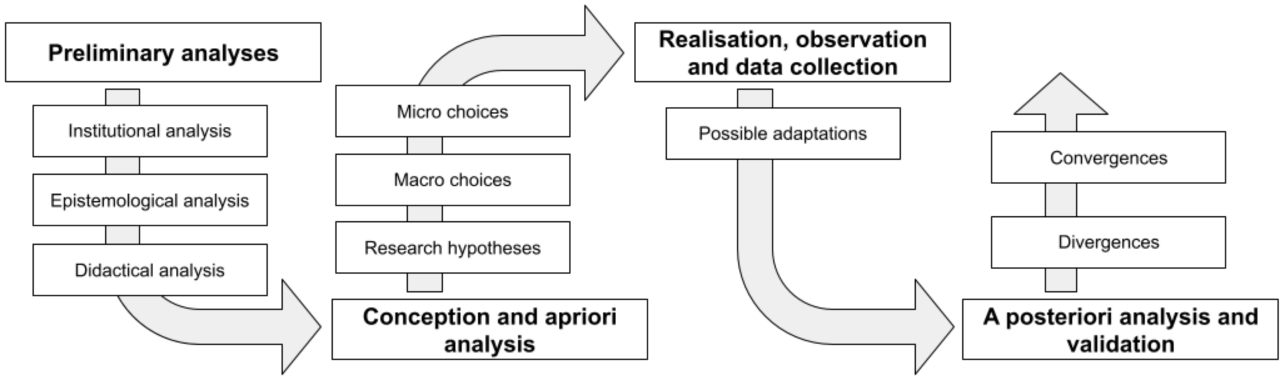


Figure 1: Overview of the phases of didactical engineering.

Each of the dimensional aspects of the four phases of DE are described in the following. The preliminary analyses should include an *epistemological analysis*, an *institutional analysis* and a *didactical analysis*. In the field of mathematics education research, epistemological analysis should focus on the mathematical content and concepts that are the subject of DE. The epistemological analysis helps the researcher establish precise goals of DE and identify epistemological obstacles that could be faced during DE. Importantly, this analysis helps the researcher in the search for mathematical situations that feature the knowledge aimed at. These situations are referred to as 'fundamental situations' in TDS (Artigue, 2015). Institutional analysis aims to identify beneficial, problematic or somehow noteworthy characteristics in the environment where DE takes place. These characteristics can be connected to students, teachers or other relevant stakeholders. The didactical analysis aims to clarify what earlier research projects have to offer DE to help guide the design. These three dimensions of preliminary analyses reflect the systematic foundation for DE as a research method. Artigue (2015) argued that institutional and epistemological analyses could benefit from a historical dimension, which can help clarify some connected constraints in DE. The didactical dimension primarily provides cognitive insight into what the research reveals about the mathematical content at stake.

The conception/design and a priori analysis phase aims to establish the research hypotheses that arise as a part of the design of didactical situations, concepts or constructs that rely on the preliminary analyses. Artigue (2015) referred to didactic variables as the macro and micro choices that are necessarily included in the design process. The macro choices are concerned with the overall didactical design, while the micro choices deal with the situational level. The a priori analysis relates the didactic variables to the research hypotheses and the preliminary analyses. These variables can, for example, be related to the characteristics of tasks, environments or resources that are presented to students as part of the DE methodology. Artigue (2015) argued that the identification of these variables leads to conjectures regarding the development of didactical situations. She further emphasised that such conjectures do not relate to individual students but "a *generic* and *epistemic* student who enters the situation with some supposed knowledge background and is ready to play the role that the situation proposes her to play" (Artigue, 2015, p. 473). This conception and a priori analysis phase creates a reference to which the realisations and data can be contrasted.

In the realisation, observation and data collection phase, the researcher should pay attention to whether the data are able to inform the goals of DE set in the a priori analysis. Artigue (2015) emphasised the importance of realisations made during this phase, as they often lead to some adaptation of the design, especially when the DE project is of significant size.

The a posteriori analysis and validation phase should be set up to contrast the a priori analysis phase. During this phase, the data collected, as well as realisations and observations made are analysed to identify convergences and divergences in relation to the preliminary analyses. The hypotheses made during the conception phase are put to the test during the validation process, which typically involves multiple data sources. Artigue (2015) mentioned that the validation process does not impose a perfect match between the two analyses and that the methods and tools for comparing the preliminary analyses and the a posteriori analyses are constantly evolving.

Additionally, I want to address the choice of DE as a research method and structure in this *kappa*. Therefore, I pose the following working question for this retrospective use of a research method. One might conclude that DE is mainly about carefully designing didactical situations from which students can learn about specific concepts. In this use of DE, I seek to frame the project through the structure provided by DE. However, I wanted to explore in what sense DE could serve as a reasonable framework for digital task design and what this view could possibly add to the existing method of DE. Therefore, I want to remind the reader to remember while reading this *kappa* to have the following question in mind:

In what sense can DE serve as an umbrella for structuring a kappa as part of a PhD dissertation?

After the concluding remarks at the end of this *kappa*, I offer reflections on the above working question.

1.6 List of paper summaries and collaborations

As mentioned in the formulation of the research *problématique*, collaboration with experts within and outside the field of mathematics education was a goal of the project. As seen in the list of papers, half of the contributions (B, C and D) were written in collaboration with established researchers within the field of mathematics education, and the other half (A, E and F) were written in collaboration with researchers outside the field of mathematics education.

I would like to briefly discuss these fortunate opportunities for collaboration. In my first change of research environment, I visited Prof. Jeremy Hodgen and Prof. Dietmar Küchemann at University College London (UCL). I learned about Prof. Küchemann's work with the CSMS[1] project and the associated tests while gathering information on ways to build proper tests for detecting lower secondary school students' mathematics-related difficulties. When assessing these CSMS tests, I found Prof. Hodgen's information, and he was kind enough to invite me to UCL to learn about task design and students' difficulties related to algebra with him and the retired Prof. Küchemann. After the change in research environment Prof. Hodgen became co-supervisor in the project.

My second change of environment was to the Technical University of Denmark (DTU), where I visited Professor at the Section for Cognitive Systems Morten Mørup. Prof. Mørup teaches a course on machine

[1] 'Concepts in Secondary Mathematics and Science' (1974–1979) was a research programme based in Chelsea College, UCL, founded by the Social Science Research Council

learning algorithms at DTU, and I contacted him to learn about the possibilities of analysing the data I had collected via MatematikFessor using machine learning. Prof. Mørup was kind enough to show interest in the project and in the data I had collected and invited me and my company supervisor Klaus Pedersen to participate in his machine learning course. Furthermore, he collaborated with me in establishing solid grounds for analysing the data using advanced statistical models.

Finally, as mentioned in the research aim, I was fortunate enough to collaborate with my fellow industrial PhD students at Edulab. Christian Hansen collaborated with me in analysing the data in the form of answers to linear equations already implemented in the online environment MatematikFessor. Lui A. Thomsen collaborated with me in designing and developing a possible intervention in the form of a teaching sequence using VR. The collaboration consisted of both the design and development of a VR application that can be used by lower secondary school students struggling to understand equation solving.

1.6.1 Summary – Paper A

This paper provides an in-depth description of the research design involving clustering students' performance solving different types of linear equations. Students' performance was clustered using data from 457,185 answers given by 37,585 students to equation tasks. Answers were distributed across 3,438 unique linear equations in a digital learning environment. The tasks consisted of different categories of linear equations. The clustering analysis contributes to the development of an online tool to provide teachers with easily accessible formative feedback. At this point, attempts to cluster students' performance have not yet been successful, meaning that no clusters have been found. Instead, a description of how the pursuit of these clusters will continue is presented alongside the research design.

1.6.2 Summary – Paper B

Despite almost half a century of research into students' difficulties in solving linear equations, these difficulties persist in mathematics classes around the world. Furthermore, the difficulties reported decades ago are the same ones that persist today. Given the immense number of dynamic online environments for mathematics teaching and learning that have emerged, we are presented with a unique opportunity to do something about this issue. This study sets out to apply the research on lower secondary school students' difficulties with equation solving to inform students' personalised learning through a specific task design in a particular dynamic online environment (matematikfessor.dk). In doing so, task design theory is applied, particularly variation theory. The final design we present consists of 11 general equation types—10 types of arithmetical equations and one type of algebraic equation—and a broad range of variations of these embedded in a potential learning-trajectory-tree structure. Aside from establishing this tree structure, the main theoretical contribution of the study and the task design we present is the detailed treatment of the category of arithmetical equations, which also involves a new distinction between simplified and non-simplified arithmetical equations.

1.6.3 Summary – Paper C

This paper presents an implementation process model for designing and implementing tasks that provide formative feedback through online learning environments used in mathematics classroom. Specifically, the model operationalises components of Vergnaud's notion of scheme. The implementation process model features a task sequence guided by controlled variation and a 'dual scheme idea'. Using such a sequence of tasks, this work illustrates how Vergnaud's notion of scheme can be used to aid teachers in hypothesising about their learners' understanding of problems involving linear equations, ultimately providing improved feedback for teachers and improved opportunities for student learning in online environments. In Denmark, the online environment matematikfessor.dk is used by approximately 80% of Danish K–9 students.

1.6.4 Summary – Paper D

This paper presents a case study of a conceptual replication study. We replicated the famous and widely cited task presented in Falkner et al. (1999): $8 + 4 = \_ + 5$. In contrast to the original study, we administered the task to the same age group (Grade 6) in a different system (Denmark) via a large-scale online learning environment, with a larger sample and two decades later. Our results indicated that the Danish students performed significantly better than the students in the original study. We discuss why this is the case and argue that online learning environments, such as the one we used, provide an important opportunity to replicate, and thus better understand, similar results.

1.6.5 Summary – Paper E

This study is concerned with establishing a means to generate better methods for analysing and learning about task design for digital learning environments. Specifically, we utilise data consisting of over 2 million unique answers from, MatematikFessor, to solve 892 unique tasks involving solving linear equations. Utilising the Multinomial Infinite Relational Model (MIRM), which could account for extensive didactical coding of the five most popular answers to each of the 892 tasks, we successfully co-clustered students and tasks into groups for further analysis. The results showed that the analysis of these clusters of tasks can provide access to valuable information on the difficulties students in the four respective groups face and what kinds of specific tasks and knowledge pertinent to what types of tasks actually cause students problems or difficulties to be anticipated by the task designer.

1.6.6 Summary – Paper F

This article presents the theoretical considerations leading to the design and development of a digital experience for teaching linear equations using a modified balance model for equation solving. We modified the balance to alter physical behaviour in a VR experience, to strengthen students' schemes for solving linear equations and to help students adapt their schemes to situations where negative numbers and mathematical negativity make equations abstract. We used the VR application in a small teaching experience with 10 students and their mathematics teacher from a Danish Grade 7 class (13–14 years of age). This exploratory study aimed to analyse and evaluate the effects of teaching with the modified balance in the VR application, which constitutes a novel teaching experience. The findings showed positive prospects for the use of VR in teaching linear equation solving, including a new equation-solving strategy enabled by the virtual environment. A majority of students gave a positive affective response to the experience, referred to and were able to apply ideas from the VR experience to linear equation-solving exercises on post-experience pen-and-paper exercises. Moreover, a particular student showed interesting behaviour and reasoning, for which we provide in-depth analysis to understand future possibilities of teaching equation solving with VR.

Chapter 2 Preliminary analyses

In this chapter, I address the first phase of DE: the preliminary analyses. I engaged in preliminary analyses appropriate to the aims and the overall research questions of the research project and in relation to the professional circumstances the project should adapt to (i.e., the difficulties the DE will encounter) (Artigue, 2015). In addition, I lay out the theoretical considerations that were chosen as a prerequisite or that helped enable and carry out the respective analyses. In the preliminary analyses, I wanted to include the three mentioned main dimensions that involve institutional analysis, including an analysis of the institutional circumstances and constraints DE will encounter, an epistemological analysis of the mathematical content at stake and a didactical analysis, including an analysis of what educational research has to offer to support the design. I attempt to address these three dimensions and relate them to the overall aim before engaging in how the included papers support these analyses and the research hypotheses.

As mentioned in the introduction, this project aimed to explore possibilities in utilising data from an online environment to improve feedback for teachers about their learners' difficulties. Before embarking on the three analyses, I want to emphasise that the didactical analysis serves as the primary analysis informing the later design process, while the institutional and the epistemological analyses primarily serve to frame the overall DE methodology used.

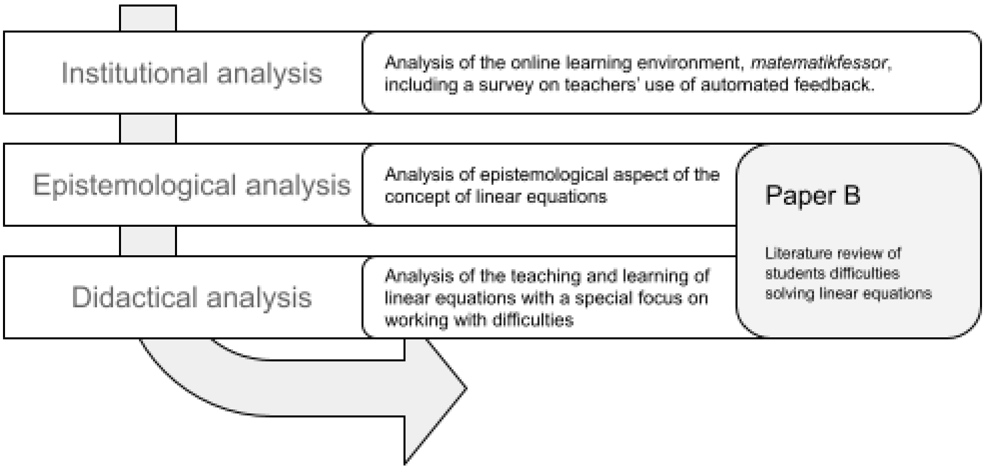


Figure 2: Structure and included papers' contributions and relation to preliminary analyses.

The preliminary analyses were structured in the order presented in Figure 2. First, I present the institutional analysis, framing the idea of working with and analysing work done in MatematikFessor. The epistemological analysis is supported by the findings from the literature regarding students' difficulties in solving linear equations. Finally, I present the didactical analysis, where I analyse what is meant by working with difficulties related to the teaching and learning of the concept of linear equations.

2.1 Institutional analysis

In this section, I explore and identify important and simultaneously problematic and beneficial characteristics of MatematikFessor. These important characteristics are, at all times, related to the overall research aim and the posed research questions to explore digital task design in the teaching and learning of linear equations in MatematikFessor. Artigue (2015) argued that institutional analysis could also benefit from a historical dimension, which can help clarify some connected constraints in DE. Based on this consideration, I attempt to provide the reader with extended insight into MatematikFessor. My reasoning is that the reader would benefit from more knowledge of MatematikFessor because, when referring to information from general digital mathematics sources gathered from research, my assumption is that the reader would then be able to contextualise the relevance of these sources to MatematikFessor and its user base.

Matematikfessor is an online learning environment (a digital learning environment presented through a website) developed by Edulab, and it is aimed at providing Grades K–12 students in Danish schools with a variety of mathematical content, both for students to explore on their own and for teachers to use in their teaching in the classroom and for assigned homework. Access to the environment is provided through subscriptions that schools purchase from Edulab. The online environment has existed since 2010, and in three years, MatematikFessor became the most used digital resource in the Danish school system.

Initially, MatematikFessor subscriptions were sold to schools based on the premise that teachers could save time from not having to correct students' assignments involving typical training exercises because the environment itself would be able to correct assignments for them. Additionally, teachers are able to access some feedback on students in the form of simple statistics based on their class performance both live and in the dashboard format. The 'dashboard' refers to the idea of a tight overview based on simple statistical summaries of students' completion of tasks and engagement with the environment. Furthermore, MatematikFessor was and is still independent from any analogue teaching materials. The initial product also featured gamified experiences, such as point systems for handing in assignments on time and for correct answers. In 2010, computers and digital devices were present in Danish schools, but not in the way they are today. The first iPad was released in Denmark in 2010 and was therefore also brand new to Danish schools. During the early years of MatematikFessor's development, in 2013, Edulab hired two teachers as a first attempt to include more of the teachers' perspective into the development of the content for the platform. Up until then, the development of the educational content had been done solely by the founder of the company Kasper Holst Hansen. This gave rise to the platform's many facets that it brings its users today. These two teachers served both as didactic support for the founder in his development of new video lectures and as a creative resource in the development of new design features. These features included the introduction of GeoGebra[2] applet-based tasks, class activity proposals, year plans and a so-called 'book case' that held what could be considered topic-specific learning trajectories.

In the short span of a couple of years, the Danish school system changed from digital sources having basically no share of the market to MatematikFessor being used by 25% of Danish schools. In 2016, the digital vs. analogue resource market for education in Denmark featured a roughly 50–50 split. Today, about 80% of schools in Denmark have a subscription to MatematikFessor.

Edulab and MatematikFessor have received plenty of criticism from higher education institutions, such as teacher education and research institutions. They have claimed that the approach inspired by online learning environments, such as MatematikFessor, have led to skill-based teaching and thereby a poor understanding of mathematics.

The input types offered by MatematikFessor, as in many other online learning environments, are limited, and in the case of MatematikFessor, only two types are available: *input fields* (the user can input a number in an empty space) and *multiple choice* (the number of options vary from 4 to 6). Therefore, the room for gathering information on students becomes rather limited in this way. Users have been able to access tasks made with an embedded GeoGebra applet for several years; however, the possibilities for data collection are not as straightforward as with the other input types.

As promised, MatematikFessor can score the answers students input immediately. However, teachers do not have easy access to students' reasoning or thinking to interpret students' answers to a degree where they are

[2] GeoGebra (https://www.geogebra.org/) is a dynamic geometry environment and a computer algebra system that is free to use.

able to provide sufficient feedback on behalf of MatematikFessor other than ‘correct’ or ‘incorrect’. The primary feedback students receive is whether the answer they have provided is correct. If the answer is incorrect, students receive additional feedback explaining how a possible path to a correct answer could look.

As a result, most tasks in online learning environments are more or less closed. Inferring students’ mathematical understanding from such tasks is therefore difficult because the environment only presents an answer, correct or incorrect, and a correct answer may be the result of incorrect, or only partially correct, reasoning, and vice versa. This becomes a significant constraint in task design, particularly when working with algebraic expressions, since it is not possible to prompt students to answer algebraic expressions without presenting multiple-choice options. Distractors as viable options among the possible multiple-choice answers are preferably chosen based on scientific or experimental findings that qualify the option as a good distractor because it indicates a particular difficulty or well-known misunderstanding. However, even good distractors have limitations, since it is usually not possible to probe student thinking and thus infer a student’s reasoning behind a particular response. Having limited options as answers to tasks might also create different solution models for students, since the answer they wanted is not present among the options. Similarly, a set of options might cause students to apply a multiple-choice solution strategy, such as ruling out options or testing the options in looking for the right answer.

In a more traditional classroom setting where students solve math tasks copied from textbooks and present the teacher with the processes they used to derive solutions through written assignments, teachers can set themselves up to acquire important assessment information about their students’ knowledge or understanding of certain mathematical concepts. This issue is related to the *problématique* and the possibilities of establishing means for fruitful formative assessments through online learning environments. Online learning environments do not immediately present teachers with similar opportunities, since teachers and the system can mostly only assess students based on the answers they provide and not the processes they used.

In the following section, I dive deeper into the possibilities for MatematikFessor to provide teachers with feedback that they can access as users.

2.1.1 Feedback in MatematikFessor

As part of the institutional analysis, I analyse and discuss the various forms of feedback the teachers and students can access when working with MatematikFessor. In the environment, teachers assign tasks for students to engage with in the classroom or at home. These tasks can be more or less handpicked by the teacher. This means that teachers can assign tasks to students in the form of an activity in which the environment itself chooses the tasks.

When students provide answers to the tasks they are assigned, teachers can receive four main types of feedback made available by MatematikFessor. The forms of feedback are centred on a statistical approach based on the correctness of the answers the students provide and are as follows:

- Live statistics
- Student/task statistics
- Topic statistics
- Complete/school-to-home statistics

Live statistics is a tool where teachers can observe their students’ engagement with the proposed tasks live. The idea behind this tool is that teachers can catch students who are either working at an unnatural pace or are continuously providing answers indicating that the tasks are either too easy or too difficult. Student/task statistics is a tool for teachers to review their students’ performance on their homework. Teachers can either get an overview of students' performance from a student or task perspective. From the task perspective, teachers can review a bar indicating how many correct and incorrect answers their class provided collectively. With a few button presses, teachers can review the actual answers their class or group provided. The topic statistics tool is intended to provide an overview of the performance of the class or group across the variety of mathematical topics the group or class has engaged with through tasks. Within a series of tasks assigned as homework, some tasks belong to a variety of mathematical topics. This tool is intended to provide a broader overview of the performance of the class or another self-defined group over longer periods of time. The topic statistics tool presents a statistical overview divided into mathematical topics, such as *area*, *multiplication* or *fractions and calculating with fractions*. These topics are not grade specific. The collected or school-to-home statistics tool aims to provide teachers with a comprehensive statistical overview of the performance of a single student over longer periods of time. This tool provides information on a student's performance within the variety of topics that the student has engaged with through tasks. Additionally, the tool provides information on how many tasks students have engaged with voluntarily and how many points, trophies or medals (digital valuables in the environment) the student has collected over the time period.

Teachers in Denmark with access to the content in MatematikFessor use these statistical overviews in various ways, not necessarily in line with the intentions behind their design. However, this does not mean that the teachers do not find the statistical overviews helpful. In the following, I present a small survey study about teachers' use of statistical overviews in MatematikFessor. The survey was conducted using Google forms, and the questions were only meant to provide an idea of the extent to which the statistical overviews are used and what they are used for.

At the end of the survey, I added an opportunity for the participants to add a comment in one or both of two sections with the prompts, 'Please add a comment about your experience if you use one or more of the statistical tools' and 'Please add a comment about your experience if you do not use one or more of the statistical tools'.

Table 1: Scores from the survey on teachers' use of statistical tools in in MatematikFessor, N = 29.

	Often	Sometimes	Never or almost never
Use of MatematikFessor in teaching	62%	35%	3%

Statistical tool	Yes	No	Did not know the tool existed
Live	72%	24%	3%
Student/task	66%	20%	14%
Topic	48%	38%	14%
Complete/school-to-home	48%	42%	10%

Unfortunately, I was not able to engage with a particularly large population of teachers for this survey. Due to circumstances within the host company, we had to cut the survey short, concluding with only 29 responses. However, I did manage to get some additional comments from teachers on the use of the statistical tools in MatematikFessor. Table 1 presents the scores from the survey distributed by the host company to what is referred to as 'the fessor panel' ('fessor' is short for professor and is often used as an abbreviation for MatematikFessor). More than 95% of the participants used MatematikFessor in their teaching, and more than 50% used it often. The live statistics and student/task statistics were more frequently used, but the remainder of the statistical tools were also used as well. Another important thing to notice is the number of participants who did not know the tools existed.

Only 77% of the teachers who declared that they often use MatematikFessor also used the live statistics tool, and the remaining 23% did not use it. Only 3 of 18 teachers did not use the student/task statistics tool.

Interestingly, 2 of 18 teachers that often used MatematikFessor did not know about the student/task statistics tool. Regarding the topic statistics and the collected or school-to-home statistics, I observed that only 10 of 18 teachers who used MatematikFessor often used these tools.

Through the survey, I identified some generalities in the comments from the participants. Most of the participants who offered comments about not using the statistical tools struggled with accessing the tools via the interface. Some participants found the tools to useless or struggled to make sense of them in relation to their teaching. A teacher who answered ‘yes’ to using all the statistical tools offered a different perspective:

> I use them [the tools] to keep up with what the students are doing. I also use them as an outset for our school/home conversations. At our school, we exclusively use MatematikFessor as a system, as we teach theme-based. I also use the tools to keep an eye on whether the students did their homework.

From this rather extreme case, we can perhaps see the power of using digital learning resources as a surveillance method. My fear is that propositions such as ‘Student X is not good at solving equations’ or ‘Student Y was 50% correct when solving tasks about finding the area’ emerge from interpretations based on the use of statistical tools as they are operating at the time of writing. Another teacher who also answered ‘yes’ to using all the statistical tools offered the following comment:

> I think the school/home statistics are not very helpful. But I use the live statistics and student or task statistics every time I assign ‘homework’ to my students. It’s a manageable way to catch uniform errors, or see if students are working, etc.

Here, we see a teacher who supposedly manages to interpret something about the errors their students make. Another teacher who often uses MatematikFessor and most of the feedback tools it provides gave a different perspective:

> Unfortunately, it takes a long time and many clicks to get around the various statistical tools. It is a pity that the topic statistics are not more precise about which grade level/topic the students can or cannot do. There is a VERY big difference between [measuring or calculating] volume in 4th grade and in 7th grade, and since the difference is huge in each grade, it would be a GREAT help to have more precise topic distinctions.

Interestingly, this teacher commented on the fact that correct answers are not tied to a specific ‘level’ of content or tasks. If a correct answer always counts as the same or represents the same value, the overall score related to topic statistics can mislead teachers if the statistics are interpreted in categories that are too broad. Lastly, I received comments such as, “They [the statistical tools] are useful but can be complex to navigate and interpret”.

These statements reinforce the research findings on feedback presented to teachers in the form of dashboards with statistical overviews (Utterberg Modén, 2021). Teachers are presented with statistics that are easily automatised in digital environments. This last statement tells me that teachers are confused about the application of these statistical overviews. If the tools are in fact useful, what are they useful for and what makes them difficult to interpret? Additional statements included the following: “I think it gives a good overview of gaps in the students’ skill” and “It helps students to ‘get moving’ and helps me to find student gaps”.

These teachers claimed that the tools provide a useful overview of where students have gaps in their knowledge. If the teachers are in fact referring to the overview and not their own interpretations of their students’ answers to tasks, learning could be reduced to the idea of what tasks students can or cannot do. One teacher said, “I tell children and parents that the target is 90% correct when working on MatematikFessor”.

The use of the statistical tools on MatematikFessor could result in the philosophy that understanding and correct answers go hand in hand to the point where a certain proportion of correct answers means that students are not experiencing relevant difficulties related to the topic they are engaged with. Some teachers found that these statistical tools on MatematikFessor benefit their classrooms, and others confirmed that they are difficult to interpret and integrate in their teaching. MatematikFessor has not yet explored the design space beyond the presentation of simple statistics based solely on correct/incorrect answers within topic areas of the offered material.

2.1.2 Common goals for mathematics in Denmark

In this section, I analyse the curricular goals set by the Danish Ministry of Children and Education in relation to the teaching of the concept of linear equations in lower secondary schools. In Denmark, mathematics education is built on a set of common goals for each subject in the mandatory school system. These common goals are reference points for what students should know or have been familiarised with after the 3rd, 6th and 9th grades, respectively. The common goals are set within four major subject fields: Danish mathematical competencies, numbers and algebra, geometry and measurement and finally probability and statistics. The Danish competencies are a set of mathematical proficiencies concerned with engagement in various mathematical situations and contexts (Niss & Højgaard, 2019). Originally, there were eight competencies; however, the Ministry of Children and Education decided to merge a couple of them into one, resulting in a new total of six competencies. The notion of competency and its relation to curricula is not the focus of this project.

The student group studied in the present project was primarily lower secondary school students (7th–9th grade in Denmark). Therefore, I mainly focus on the guidelines for this age group. In the common goals, specific sections are dedicated to equations a student should be able to comprehend after 9th grade. The following are my translations of the common goals for knowledge about equations after 9th grade (Ministry of Children and Education, 2019a, p. 8, my own translation).

- The student is capable of developing methods for solving equations.
- The student possesses knowledge about strategies for solving equations.
- The student can pose and solve equations and simple inequalities.
- The student possesses knowledge about equation solving with and without digital tools.
- The student can pose and solve simple systems of equations.
- The student possesses knowledge about providing graphical solutions for simple systems of equations.

There are no guidelines specific to the concept of equations in the set of goals related to students' knowledge after the 6th grade. However, there are a couple of points where the word 'equation' is mentioned (Ministry of Children and Education, 2019a):

- The student can come up with solutions for simple equations using informal strategies.
- The student possesses knowledge about the meaning of the equals sign and about informal strategies for solving simple equations.

To find out more about what the prescribed common goals specifically mean when stating that the students should acquire knowledge about the 'meaning of the equals sign', we have to dig further into the lesson plans for Danish primary and lower secondary schools. In the lesson plan, we are provided the following description.

> Teaching takes place on the basis of the students' intuition, which is based on the use of concrete materials, drawings and own notes, 'guess and try', as well as explanations in everyday language. Teaching can include, among other things, problems and calculations

> from everyday life that can be described with equations. Students must develop an understanding that the equals sign means that the expressions on the left and right sides of it have (or must have) the same value (as opposed to an understanding that the equals sign is a signal that calculations must take place), a concept that is central to working with equations. (Ministry of Children and Education, 2019a, p. 23, my own translation)

This text reveals that in the prescribed lesson plan and common goals for mathematics in Denmark, the concept of equations is well represented, and a focus is on strategies for solving equations and the interpretation of the equals sign. The wording connected to the interpretation of the equals sign is not far from that found in international literature (Kieran, 1981; Matthews et al., 2012; Prediger, 2010).

From the prescribed lesson plan, we get a small but important distinction between the terms 'unknown' and 'variable'. Even though the terms could be confused when working with equations as opposed to functions, the explanation of how students should learn to replace generalised numbers with letters is present. Learning about different number sets is loosely mentioned throughout the legal texts. Knowledge about negative numbers is, however, referenced as a focus point from the 4th through 6th grades (Ministry of Children and Education, 2019b).

The idea of including the prescribed curricula is to help inform the design in the next phase of the DE. From an institutional perspective, on the curricular side, there is a focus on learning about strategies for solving equations, and there is a focus on the shift in meaning of the equals sign coming from an operational view of the equals sign as relational. The comparison of the opportunities for feedback for teachers in MatematikFessor with the idea of learning about equations from a ministerial point of view in Denmark shows that these ideas do not seem to coincide. Feedback based on the presentation of correct answers might require that the tasks are of a certain quality to specify something about the strategy used or knowledge applied. In the following sections, I present the epistemological analysis, covering the idea of what it means to know about linear equations and equation solving from a lower secondary school perspective.

2.2 Epistemological analysis

In this section, I want to dive into the epistemological and historical perspective on what linear equations are and have been. Equations, and especially linear equations, have been a part of mathematical problem solving since the Babylonians (Høyrup, 1998). The mathematical field of algebra has been the domain for introducing generality to mathematical expressions through the introduction of letters as placeholders for mathematical structures (not always numbers), satisfying appropriate conditions. Several research studies summarised students' difficulties in learning the elements of algebra (e.g., Kieran, 2007; Rhine et al., 2018), which testify to the relevance of an epistemological analysis of the concept of linear equations. What does it mean to know about linear equations, and how are algebraic notation and the algebraic thought process relevant to learning about linear equations? As a part of the epistemological analysis, I refer to the findings from Paper B, which presents a review of the literature concerning students' difficulties learning the concept of linear equations and equation solving.

Through the following sections, I aim to confine the concept of (linear) equations to accommodate the elements of the institutional analysis and the overall aim of the project. To conduct an appropriate epistemological analysis of students in lower secondary school, I make some simplifications to the concept of (linear) equations that I return to in the discussion and implications of the preliminary analyses after the didactical analysis.

2.2.1 What is an equation?

An equation is a mathematical proposition that two expressions are related to each other via the equivalence relation called 'equality'. Equivalence relations are not specified or defined in great detail here, since it is not important for the present study. In general, we refer to the two expressions separated by the equality symbol (the equals sign) as the *left side* and the *right side.* It should be noted that it does not affect the proposition if the two expressions are switched as per the definition of the equals sign as an equivalence relation. This proposition, when we discuss equations, can therefore be logically true or false, since equivalence can be considered met or not met. The proposition is considered false in situations where equivalence cannot be met. However, in modern textbook problem solving (tasks), there will usually exist one or more conditions that ensure that equivalence can be met. In particular, when talking about equations in the simpler sense of the term, a set of unknown values in one or both of the two expressions allows the proposition to become true. Sticking with this simpler definition of an equation, we can define general simple linear equations, which are relevant to this project and suited for this epistemological analysis as propositions of the structure:

$$a_1x \pm \cdots \pm a_nx \pm b_1 \pm \cdots \pm b_n = c_1x \pm \cdots \pm c_nx \pm d_1 \pm \cdots \pm d_n$$

$$a_1 \dots a_n, b_1 \dots b_n, c_1 \dots c_n, d_1 \dots d_n, x \in Q, n \in N.$$

Historically, equations were not written using the symbols (both operator and letters as unknown, known or general numbers) but were written in plain text. Diophantus (born around 200 AD) was the first person known to use algebraic symbolism when solving equations (Katz, 2009). Herscovics and Kieran (1980) provide a perhaps more precise or relevant definition of what an equation is in this context: "an equation as an arithmetic identity with a hidden number" (p. 575).

2.2.2 What is a solution to an equation?

To satisfy the equivalence relation, we talk about an equation as having a set of solutions. The cardinality of this set is either 0, 1 or infinity for linear equations equations with a single unknown. In this context, we consider equations with only a single unknown value, x, following the general formula from Section 1.1.1. This unknown value belongs to a set of solutions. Equivalence will be satisfied if and only if the unknown value belongs to the set of solutions. There may be a given starting condition that the set of solutions must satisfy. However, this idea belongs to situations in which the equation is considered part of a context or a system of equations. In the context of this study, we only consider equations with a single solution, and this solution belongs to the set of rational numbers. The solution to the generalised linear equations is as follows:

$$x = \frac{d_1 \pm \cdots \pm d_n - (b_1 \pm \cdots \pm b_n)}{a_1 \pm \cdots \pm a_n - (c_1 \pm \cdots \pm c_n)}$$

In the simplified version, the solution would be expressed as,

$$x = \frac{d - b}{a - c}$$

Historically, different methods have been used to solve equations. For example, the Babylonians would solve systems of linear equations by false position, beginning with an assumption that is soon altered into the solution (Katz, 2009). In recent history, we began solving equations by using several equation-solving strategies that can each be valuable in different situations (Linsell, 2009a). Such strategies might, for example, include the ‘trial and error’ strategy where the solver can make deductions by replacing the unknown with a number and determining whether the number is too large or too small before attempting with a better number. In some cases, one could simply ‘know’ the solution based on perhaps more basic knowledge about multiplication tables or additive structures. In Archimedes’ work from more than 2000 years ago, we see the strategies of the construction of the idea or working backwards when solving equations to unwind the problem to come to a solution (Katz, 2009).

2.2.3 What is an unknown?

When we talk about an equation in its mathematical sense, I need to emphasise what is meant by the term ‘unknown’ or ‘an unknown’. One could easily be confused by the terminology when discussing the unknown magnitude or value of an equation. The words ‘unknown’ and ‘variable’ are occasionally interchangeable for the same elements in mathematics, since equations can been viewed as functions evaluated at a certain point. I have also not been able to acquire any other evidence that constitutes a written definition of the correct use of the terms ‘unknown’ and ‘variable’. An unknown size or number in an equation is not necessarily the value that is searched for or in any other way constitutes the solution, since other values (such as *a* or *b*, serving as independent terms or coefficients) might also not be known. For that reason, it becomes of utmost importance that in the mathematical context that what constitutes the unknown value is well defined and uncovered to establish a set of solutions. In other words, the word ‘unknown’ can cover all sizes that are not known elements in an equation of equivalence. The word ‘unknown’ is here used as Prof. Niss (personal conversation) would refer to it for its unknown value that we are ‘in search of’ in order to satisfy the proposition of upholding equality between two expressions.

2.2.4 Epistemological obstacles

In this section, I attempt to concretise important (relevant) concepts that serve as prerequisites for knowing about linear equations. This should not be considered an exhaustive list; rather, it should be seen as a healthy analysis. Kieran (2007) argued that our ability to successfully manipulate the symbols of algebra requires that we are familiar with the properties of mathematical operations and relations. Furthermore, recognising the abstract ideas hidden behind the symbols allows for knowing which transformations are legal. Drawing on Bachelard (1938), Brousseau defined epistemological obstacles as:

> Forms of knowledge that have been relevant and successful in particular context, including often school contexts, but that at some moment became false or simply inadequate, and whose traces can be found in the historical development of the domain itself. (Artigue et al., 2014, p. 49)

Dissecting the concept of linear equations in terms of epistemological obstacles, I wish to analyse what an equation is and what is meant by a solution to an equation and ultimately what could be meant by knowing the concept of linear equations. In the didactical analysis, I attempt to establish what is meant by experiencing difficulties related to not knowing about equations and their solutions.

I wanted to establish what an equation is and what is meant by a solution to it from both a mathematical and a historical perspective. I begin with the historical perspective, since we can acquire a lot of meaning from history when discussing the aspect of epistemological obstacles. Brousseau (1997) tied the meaning of epistemological obstacles to the development of mathematics through its history. Epistemological obstacles are not to be avoided when learning mathematics, and learning is very tightly coupled with overcoming epistemological obstacles for oneself. What ties epistemological obstacles to the history of mathematics is that these obstacles can often be found in the history of the concepts themselves (Brousseau, 1997). I am not emphasising that students and teachers should delve into replicas of the situations from the history of mathematics but that they should respect the epistemological value in overcoming the obstacles in these situations.

Many researchers agree that embarking on a journey to solve problems in algebra requires a more mature understanding of the concept of equality compared to what is expected from an understanding necessary for the signalling that a directional transformation has taken place (Kieran, 1981; Prediger, 2010). By directional transformation, I mean when the equals sign is used to separate what, for instance, is added and what the addition sums to. Something about the directional aspect of equality stemming from reading a text results in a convention that the 'answer' is on the right and whatever is on the left accumulates to or generates this answer. Additionally, we encounter situations where the equals sign is already present in a situation signalling that some calculations should take place.

In a review of the literature, I found that a significant source of difficulties related to learning the concept of linear equations is related to negative numbers. For many years, these numbers have been disregarded because they intuitively and historically did not make sense. In the history of mathematics, we are familiarised with the fact that negative numbers were first accepted for incorporation in calculation and manipulation in the beginning but not as results or solutions (Katz, 2009).

In his work from 1557, Robert Recorde invented the equality symbol or the equals sign depicting two parallel lines as we know it today. As Recorde wrote, "no two things can be more equal" (Katz, 2009). Before the year 1557, mathematicians would simply write out the phrase "is equal to" instead of using the symbol. The difference between the symbol and the phrase might not be what is causing students difficulties; rather, the shift in meaning as students advance through middle and lower secondary school is the source of confusion (Kieran, 1981). Reading the phrase "*a* is equal to *b*" does not immediately give rise to the symmetrical idea of "*b* is equal to *a*" which is otherwise part of the definition of an equivalence relation. Without reflecting too much about this notion, I think it is important to realise that in large parts of the world children are taught to read from left to right and therefore, reading mathematical expressions containing an equals sign symmetrically might not be obvious.

In the literature review in Paper B, which contributes significantly to the epistemological analysis, and in these sections about the epistemology of linear equations, I have discussed what an equations is and what is meant by a solution to it. In the following sections, I continue the preliminary analyses with the didactical analysis covering the teaching and learning of the concept of linear equations before summarising and discussing the implications of the preliminary analyses.

2.3 Didactical analysis

In this section, I aim to establish what prior research projects have to offer to help guide the research design in a didactical analysis. This analysis provides cognitive insight into what prior research discovered about the teaching and learning of linear equations. First, I present the different parts covering this didactical analysis. In the literature review, we established an analysis in the form of a review of the difficulties related to the learning of and to working with linear equations. Additionally, I want to establish an analysis that enables the emergence of possible fundamental situations (Brousseau, 1997) in online learning environments through this DE and this project:

> A situation is itself a system, 'the set of circumstances in which the student finds herself, the relationships that unify her with her milieu, the set of "givens" that characterize an action or an evolution'. (Brousseau 1997, p. 214)

According to TDS, didactical situations are when a teacher arranges the devolution of an appropriate adidactical to students. The adidactical situation is one or more situations that characterise the intended target knowledge (Brousseau, 1997): "Each item of knowledge can be characterized by a (or some) adidactical situation(s) which preserve(s) meaning; we shall call this a fundamental situation" (p. 30).

In TDS, a didactical situation is therefore a situation designed to facilitate teaching and most of all learning (Brousseau, 1997). According to the overall aim of the project and DE, I attempt to enable myself to analyse and design opportunities (i.e., didactical situations) for the teaching and learning of linear equations in online learning environments relevant to lower secondary school students. In other words, I am specifically interested in the tasks and task design in online learning environments in relation to the teaching and learning of linear equations. The didactical design and mathematical problems (situations) should be appropriate for students in lower secondary school (in Denmark). Furthermore, these situations (tasks) should, according to the overall research aim, address epistemological obstacles mentioned in the epistemological analysis.

Furthermore, Brousseau claimed that the identification of such epistemological obstacles is essential to establishing and analysing didactical situations. Brousseau (1997) argued that to do mathematics, students should be asking themselves questions and solving problems. Through this section, I thus also attempt to analyse what it means to be solving problems with linear equations.

To enable a research-based foundation to create and identify epistemological obstacles related to the teaching and learning of linear equations in lower secondary school, I reviewed the literature to find out what research studies have to offer in the identification of general epistemological obstacles. Initially, I imagined that such a literature study would and should result in a comprehensive list or a categorisation of misconceptions students had or errors students made when solving linear equations. I soon came to realise that the terminology referring to phenomena leading to students' epistemological obstacles in working with linear equations was not, in my opinion, comprehensive enough. One might exemplify this issue via the distinction between what errors students make while working with or solving linear equations and the reason for or the origin of behaviours resulting in these errors. I conducted a hermeneutic literature review (Boell & Cecez-Kecmanovic, 2014), presented in Paper B, of the difficulties related to the learning and solving of linear equations. However, I did not explicitly search the literature for epistemological obstacles related to the teaching and learning of linear equations because of the terminological inconsistency related to research on students' difficulties in learning the concept of linear equations. Therefore, I came to expect something a little different from the literature review I conducted in comparison to the initial aim.

Instead of a comprehensive list of misconceptions or categorisations of such, I imagined that knowledge about situations where students might find themselves lacking sufficient knowledge (situations touching upon epistemological obstacles) to identify how to interpret or proceed to solve linear equations could be beneficial. In addition, I wanted to come closer to understanding why these situations were causing students to experience difficulties. In other words, what had the students not learned or what knowledge had been applied incorrectly, hindering their accomplishment of what was expected from them in the situation? Discovering this knowledge should enable the design of didactical situations that record students' epistemological obstacles.

In the following, I dive a bit further into some important terminology that naturally became a part of my work addressing students' knowledge or proof of absent knowledge. The term 'misconception' refers to a faulty conception, a piece of knowledge or understanding that is, in this context, mathematically incorrect. To some extent, this view of students' understanding quickly becomes insufficient from a didactical design perspective. Here, I shall draw on the idea of the concept image and concept definition (Tall & Vinner, 1981). Throughout a student's time in school, the idea of their concept image might change due to instruction and experiences in mathematics-related situations. However, the concept definition should also be considered dynamic. This can be exemplified when introducing the extension of number sets to the meaning or definition of a concept. For example, we might look at subtraction. When working with only the natural numbers, subtraction is not a well-defined operation, since the natural numbers are not closed under subtraction. A student might be quite familiar with subtraction but hold a conflicted or situated understanding of the numbers and qualities of different numbers sets. All this is to establish that a concept definition is also a dynamic phenomenon. This is why I avoid the use of the term 'misconception', since I do not expect students to hold a conception or a concept image that is either specifically true or false in every situation but rather true or false depending on the situation. What is much more interesting to observe is the situated reason for erroneous answers or mathematical actions. Prof. Küchemann wrote in his dissertation the following:

> A second objective of the CSMS research was to identify the difficulties that students encountered in mathematics and to indicate reasons for these difficulties. Of particular interest were systematic errors or misconceptions that had some logical, if flawed, basis, often the result of an overgeneralization of a 'correct' mathematical idea. A deliberate effort was made to develop items that provoked common wrong responses of a sort that could be interpreted in terms of an underlying strategy. (Küchemann, 1981, p. 30)

I realise that this might come down to semantics and that the terms 'misconception', 'understanding' and 'knowledge' hold different meanings for researchers (Smith III et al., 1994). What I want to clarify is that in this project inspired by the Maths Counsellor Programme, I am concerned with equations with possible diagnostic value. Therefore, I am very much interested in the reason behind a wrong answer given to a task in an online environment, especially the concept of students experiencing situated difficulties. I made the decision to use the term 'difficulty' in reference to Jankvist and Niss (2015). Drawing on Kieran (2007), they specifically presented two categories of difficulties related to understanding the concept of equations:

> The first kind of difficulty ... is to do with goal-oriented transformation of equations (and, more fundamentally, algebraic expressions) into equivalent ones by way of permissible operations. [...] The second kind of difficulty, which appears to be of a more fundamental nature, is to do with what an equation actually is, and with what is meant by a solution to it. (Jankvist & Niss, 2015, p. 276)

With this adaptation, I imagined that I would be able to establish to a higher degree a clear reference to what I mean by students' situational behaviour resulting in an erroneous answer to a linear equation. In a later section (3.1), I shall dive deeper into a more nuanced perspective on situated difficulties and actions.

2.4 Discussion and implications of the preliminary analyses

In this section, I explain the implications of the preliminary analysis and how they feed into the following phases of the DE and the publications that serve as the contributions of this PhD dissertation. In addition, I want to establish how these analyses will affect the hypotheses and the design in the following sections. I wanted to conduct a literature review that would enable me to analyse, design and develop tasks and teaching strategies related to the learning of linear equations among lower secondary school students, preferably with a link to online learning environments. Additionally, I wanted to establish a state of the art of known difficulties experienced by lower secondary school students related to working with linear equations.

An important point that I want the reader to remember is that I am working with a specific online learning environment (MatematikFessor) in mathematics education through an industrial PhD project. This is important because when drawing upon the claims and findings found in research publications from a variety of research environments and also discussing (digital) resources or tasks, one must keep in mind that these international publications are not related to the specific learning environment I am working with. Because MatematikFessor is a purely Danish digital learning environment, some of the findings made by other research projects regarding digital resources might not fit perfectly into this particular framing; instead, they might serve as different perspectives or general insight. Furthermore, I am working with an established institution and therefore did not have full control or say in what and how ideas could be implemented or taken into consideration.

The main focus of the project on its epistemological side is on linear equations and establishing an idea of fundamental situations regarding equations in online learning environments. Even though an overall aim is to establish new grounds for a new idea or approach to the provision of feedback in online learning environments, an initial focus is on epistemological obstacles or difficulties related to learning the concept of linear equations and linear equation solving. Through Paper B, I established an idea of what fundamental situations in learning the concept of linear equations and equation solving should encapsulate. This similarly informs the understanding of the difficulties related to linear equation solving experienced by lower secondary school students.

Reflecting on the terminological discrepancies regarding difficulties, I decided to change my plan regarding how I was going to move forward in discussing difficulties related to understanding the concept of linear equations. When discussing my project and the scope, since I set out to explore students' footprint through data from an online learning environment, I changed my perspective towards analysing and hypothesising about the behaviours or the actions that led to such footprints. Theorising about students' conceptions is perceptively much more difficult than hypothesising about their actions. My initial understanding of how a data footprint would be analysed led me to think about how a conception or understanding should be considered situational.

I draw on the theory of conceptual fields (TCF) (Vergnaud, 2009). Much of the reason for doing so is based on the idea of analysing the actions (and thereby the schemes driving the actions) of the students when both designing tasks for generating a basis for a digital footprint and when reviewing their digital footprints. Gerárd Vergnaud offered an extensive description of mathematical action as action guided by, to the enactor, an appropriate scheme. I found it very meaningful to conceptually replace the concept of understanding or knowledge with that of an action or the scheme. TCF serves a dual purpose:

> The theory of conceptual fields is a developmental theory. It has two aims: (1) to describe and analyse the progressive complexity, on a long- and medium-term basis, of the mathematical competences that students develop inside and outside school, and (2) to establish better connections between the operational form of knowledge, which consists in action in the physical and social world, and the predicative form of knowledge, which consists in the linguistic and symbolic expressions of this knowledge. As it deals with the progressive complexity of knowledge, the conceptual field framework is also useful to help teachers organize didactic situations and interventions, depending on both the epistemology of mathematics and a better understanding of the conceptualizing process of students. (Vergnaud, 2009, p. 83)

Through the publications included in this *kappa*, I propose that TCF is a valuable theory for both working with task design aimed at addressing epistemological obstacles (difficulties) and as a theory for assessment with diagnostic purposes. Working with the scheme as an organiser of activity, my reasoning is that the epistemological obstacles presented to students in an online learning environment when working with linear equations constitute a meaningful conceptual field related to meaning (2) given by Vergnaud.

The idea of providing feedback through MatematikFessor is not directed at allowing teachers to hypothesise about their learners' erroneous actions or about giving them the ability to improve instruction based on such hypotheses. Based on the institutional analysis, I gather that the Danish Ministry of Children and Education has several foci related to Danish students' learning about linear equations. Included among these foci are strategies for solving linear equations. With the idea of improving feedback for teachers using online learning environments in their teaching, I hope to enable teachers to hypothesise about their learners' conceptual fields in relation to linear equations, as well as their learner schemes for solving linear equations appropriate for lower secondary school.

Chapter 3 Conception and a priori analysis

The research design process following the DE principles incorporates the determination of didactic variables, which condition the interactions between students and knowledge, between students, between students and teachers and thus the opportunities that students have to learn (Artigue, 2014). In the case of this project, the main aim of the project was to investigate the improvements in the assessment and feedback generated through the analysis of students' mathematical difficulties learning the concept of linear equations and equation solving through tasks in an online learning environment. This has some implications for how hypotheses and didactic variables might be viewed and discussed. Additionally, fundamental situations might require a different interpretation, since the idea is not to learn about linear equations; rather, it is about the measurement of this learning.

In this second phase of DE, conception and a priori analysis, the research hypotheses are established. Design requires a number of choices from global to local. These choices determine the didactic variables that condition the interactions between students and between students and teachers, ultimately implicating the opportunities that students have to learn. In line with TDS, in design, particular importance is attached to the following:

- The search for fundamental situations (i.e., mathematical situations encapsulating the epistemological essence of the concepts).
- The characteristics of the milieu (the learning environments) with which the students will interact to maximise the potential it offers for autonomous action and productive feedback.
- The organisation of devolution and institutionalisation processes through which the teacher makes students accept the mathematical responsibility of solving the task and connects the knowledge they produce to the scholarly knowledge aimed at.

Inspired by the Maths Counsellor Programme's framework of detecting, diagnosing and intervening, the research hypotheses were derived from this idea. All of the papers presented through this *kappa* establishes or enables the design of didactical situations in different forms. During this phase of the DE, I lay out my a priori analyses relevant to the respective designs of the different areas, which I have touched upon in exploring the established research questions and the preliminary analyses.

Before establishing the research hypotheses that serve as catalysts for answering the posed research questions through DE, I wish to present a theoretical perspective that has helped me come to terms with the idea of measuring knowledge or difficulties in didactic situations.

3.1 My view of difficulties, knowledge and the measuring of such in relation to linear equations from a task design theory perspective

In this section, I extend the discussion that was started in the preliminary analysis with theoretical constructions to establish a theoretical standpoint for addressing difficulties related to making sense of and solving linear equations in an online learning environment. I realise that to establish such a standpoint, I need to clarify not only how the chosen theoretical construct benefits this standpoint but also provide a general view on why these theoretical constructs benefit the standpoint from a design perspective. I wish for the reader to follow me in the theoretical process of establishing didactical situations suited for online learning environments and for working with or measuring mathematics-specific difficulties related to linear equations.

Task design has a very strong link to assessment, both formatively and in a summative way, in measuring knowledge or capabilities. I take this perspective through the lens of the notion of scheme as a part of TCF (Vergnaud, 2009). Based on the idea that engaging in a (mathematical) situation triggers an action, we can gather that a didactical situation is set up to trigger specific actions. These specific actions could be seen as actions that potentially drive individuals to acquire new knowledge or to refine their knowledge according to TDS. Similarly, we can view the didactical situation as set up both in the attempt to measure knowledge (an assessment situation) and to establish new knowledge (a learning situation). Essentially, we utilise tasks or problems in both learning and assessment situations. I advocate that in the design of mathematical tasks, we could justify the presence of both intentions: tasks designed for learning situations and tasks designed for assessment situations. In many cases, tasks would be able to serve both purposes. This is followed by a discussion of how some tasks might be better suited for either learning or assessment.

From an assessment perspective, I propose the idea that knowledge of students experiencing mathematics-specific difficulties related to some mathematical concepts can be conceptualised through their mathematical actions and thereby possibly related to their scheme for engaging in that particular mathematical situation. In the following, I go into detail about the theoretical constructs that provide the framework for this view.

3.1.1 The concepts of diagnostic task and diagnostic value

In this section, I explain my reasoning for highlighting the diagnostic task, or rather the diagnostic value, the diagnostic potential and diagnostic relevance of a task. This is, of course, a matter of defining that within a situation, we can identify different roles or stakeholders in the assessment of knowledge. Therefore, within the diagnostic situation, I propose that three entities can be defined: the task, the enactor and the observer (see Figure 3).

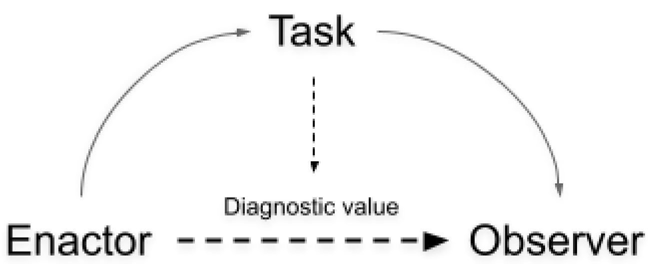


Figure 3: Structure of the diagnostic situation.

When I mention the concept of situation and, in particular, the notion of the diagnostic situation, I am again referring to the TDS. Diagnostic situations can be thought of as didactical situations in which the observer (the teacher) seeks to observe some interaction between the enactor (the student) and a task. On the notion of value, I would like to refer to philosopher DeWitt H. Parker:

> To common sense the inevitable starting point in philosophy, the universe, is divided like Gaul into three parts,-an inner world, an outer world, and a world between; and if we look for instances of value, we seem to find them in all the three parts of the universe. In the outer world they seem to reside in articles of use and consumption, like railroads and bread; in the inner world, within needs and desires; in the world between, among objects which we call beautiful. (Parker, 1929, pp. 303–304)

Additionally, Parker (1929) emphasised that value cannot exist in absolute separation from the mind, meaning that value is in fact a subjective notion. I realise that beauty might not be relevant in a discussion about diagnostic value; however, the idea about the value bearing meaning, and only to some, is important in this context. Following this idea, I argue that diagnostic value only bears meaning to the observer and only when the observer desires this piece of information. This meaning is what drives the main idea of formative assessment.

The three entities fulfil different roles. The role of the task is to enable the observer to hypothesise about the desired piece of knowledge of the enactor. The role of the observer is, in an efficient and as deterministic as possible manner, to gather information about a piece of knowledge of the enactor. The enactor is less invested in the situation by design. However, if the diagnostic situation is to be considered valuable to the observer, the enactor should easily, willingly and, to some extent, precisely display certain characteristics about the piece of knowledge the observer seeks to acquire information about.

For example, one could imagine the medical procedure of getting a diagnosis. In medicine, the term 'diagnosis' refers to the identification of the nature of an illness or other problem by examination of the symptoms. Therefore, when we are discussing the more abstract term of 'concept of knowledge', we can perhaps rely a bit on the medical term when observing a problem and attempting to identify its nature. When we adopt the term 'problem' from medicine, we are referring to a piece of problematic knowledge. This ties in well with the idea of experiencing difficulties, when students seek to apply some piece of problematic knowledge or knowledge inappropriate for the situation. To clarify, the enactor could be applying a piece of knowledge that the observer could view as problematic simply because the action of applying the piece of knowledge could be a display of incorrect mathematics. It is important that this piece of knowledge only really makes sense in its consideration as problematic from the observer's perspective, since I assume that it would be inappropriate for the enactor itself to make the diagnosis. In other cases, the action of applying the piece of knowledge might be insufficient for similar situations and therefore deemed problematic by the observer. For example, a strategy for solving a simple linear equation (such as guess and check) could be inappropriate when solving algebraic equations with the unknown present in multiple terms and therefore be considered problematic.

With this more philosophical aspect of diagnostics out of the way, we can move on to discussing the diagnostic value and potential of a task, since the task is the entity that is particularly constructed or designed in the situation.

I would like to introduce the idea of the diagnostic value of a task as a continuous value ranging from low to high. Some tasks can hold a higher diagnostic value, and some tasks might hold a lower value. Some tasks might even be considered as having a negative diagnostic value, in the sense that the task might present the observer with wrong or misleading information about certain characteristics of the enactor's piece of knowledge. The diagnostic value of a task is not always relevant to a discussion. In complex situations with no predominant purpose for the observer to gather data for assessment, the meaning of the diagnostic situation becomes irrelevant to discuss, as does the meaning of the diagnostic value of a task. However, it might be relevant to discuss the diagnostic value in smaller or isolated parts of the same situation.

It is possible to discuss determining both an a priori diagnostic value and an a posteriori diagnostic value. When establishing situations with tasks meant for diagnostic purposes, I imagine that this process includes some design principles or considerations in an attempt to secure success in the endeavour. A task can receive either a correct or an incorrect answer. In this terminology a task has one correct answer and many incorrect answers. Imagine the set of incorrect answers divided into two subsets; answers that are interpretable, i.e. corresponding to a recognised (mathematical) erroneous action, or answers that are uninterpretable, meaning that no reasonable erroneous action seems to have led to the answer. Furthermore, we need to evaluate what proportion of the empirically collected answers fall within each of the three possibilities, correct, interpretable and uninterpretable. Lastly the set of interpretable answers could be either small or large depending on the number of didactically different mathematical actions connected to the answers. I exemplify high diagnostic value in the following criteria:

1. The subset of interpretable answer should be small, meaning that some diagnostic value is lost when the set of recognised (mathematical) erroneous actions grow in size.
2. The distribution of the answers among the three possibilities must not favour the uninterpretable answers, meaning that the proportion of answers that are considered uninterpretable should be low.

When the number of didactically different interpretable answers grow, the diagnostic value drops. However, the value significantly drops when the distribution of answers starts to favour the group of uninterpretable answers. In addition, should we encounter a situation, where several recognised (mathematical) erroneous actions correspond to the same single answer, then the task is unreliable, and the diagnostic value is considered negative.

When discussing the a priori diagnostic value of a task, I refer to the term *diagnostic potential.* Hereafter, we can empirically verify the diagnostic potential culminating in the diagnostic value of the task. This is where the argument of how the complexity in the formulation of the task is important to recognise when establishing the diagnostic potential. Diagnostic value relates in its definition to diagnostic potential in that the reasons that should correspond to the answers the enactors provide are what was a priori determined in the design of tasks for diagnostic purposes. When designing tasks, a designer most likely wants to determine the diagnostic potential of a task. I argue that in most cases, task designers do this automatically. When working specifically to design tasks meant for diagnosing some particular difficulty or the nature of some particular problematic characteristics about a piece of knowledge, I argue that the more closed or to the point a task is, the easier it becomes to establish diagnostic potential. The probability of the task accidentally achieving a negative diagnostic value could increase with the level of complexity in the formulation of the task. When a task comes to complex, I refer to the idea of concepts or entities (measurable or non-measurable) that could influence what answers a task receives, the number of analytically different answers could rise and therefore also the number of uninterpretable answers. The goal of designing a task meant for holding a high diagnostic potential and ultimately value, should preferably be of a nature in which the probability that the enactor displays characteristics irrelevant to the observer (uninterpretable answers) is minimised.

When discussing the idea of diagnostic relevance, that is, the relevance of assessing the characteristics of the piece of mathematical knowledge, I imagine that for a diagnostic task to fulfil its potential, the task needs to be relevant to the relationship between the enactor and the observer. In this sense, we can discuss how fitting a diagnostic task is in the desired diagnostic situation. Discussing such relevance can be tricky, as relevance can exist on different levels, given the purpose of the diagnostic situation(s) and the different stakeholders involved. Relevance can be discussed in terms of purpose, in which case the diagnostic situations should be fair to the enactor. If the task is not relevant to the enactor's current educational level, the purpose of the diagnostic situations is skewed and becomes irrelevant as assessments. In a similar fashion, the diagnostic task should be appropriate for what the observer wishes to learn about the enactor.

When discussing the idea of diagnostic relevance in online learning environments, such as MatematikFessor, I want to discuss the relationship with the specific input types. In a best case scenario, a task with a high diagnostic value is paired with the entire set of related answers or an input field with the opportunity to input the right answer. However, if a task is not paired with the entire set of related answers, one might suspect that an important answer could be left out, hence lowering the diagnostic value. Likewise, if the answer to an equation was 1/3 and the input field did not allow for fractions or infinitely long decimal numbers, the task could lose its ability to measure what might be expected of it.

In this context, I argue that a sound preliminary analysis with the aim of securing relevance in terms of making a fair assessment of the students' ability to solve linear equations or manage different concepts when doing so is very meaningful. Working towards achieving probabilities of diagnostic relevance and diagnostic potential, I argue, is necessary when measuring the diagnostic value of tasks.

3.1.2 The theory of conceptual fields and the notion of scheme

The TCF and the notion of scheme is a significant theory in establishing my theoretical standpoint for designing didactical situations for online learning environments. Vergnaud (2009) introduced TCF as follows:

> At the same time a set of situations and a set of concepts tied together. By this, I mean that a concept's meaning does not come from one situation only but from a variety of situations and that, reciprocally, a situation cannot be analysed with one concept alone, but rather with several concepts, forming systems. (p. 86)

Vergnaud claimed that even though the definition of a conceptual field is quite clear, the boundaries or respective overlaps between them are not (Vergnaud, 1988). Vergnaud (1988) continued to clarify that conceptual fields and the definition of concepts are particularly interesting for researchers (and teachers) to interpret because they enable the understanding of the following:

> 1. Mathematical concepts are rooted in situations and problems.
> 2. We need to analyze and classify these situations and the procedures students use to deal with them. Mathematics is an indispensable tool for this analysis.
> 3. Students' ideas and competencies develop over a long period of time. Teaching students at a particular grade requires that one have a fair idea of the steps they may or may not have gone through and of the next and ultimate steps one would like them to reach.
> 4. Symbols (signifiers) do not refer directly to reality but to the cognitive components (signified) underlying students' behavioural procedures. I call the cognitive components [operational] invariants. Categories, object, properties, relationships, theorems-in-action […] are invariants. We must pay attention to the distinctions among situations, invariants and symbols. (Vergnaud, 1988, p. 142)

In addition, Vergnaud introduced the notion of a scheme. With reference to Jean Piaget and Immanuel Kant, Vergnaud (1998) described a scheme as functional dynamic totality and "a universal that is efficient for a whole range of situations, and it can generate different sequences of action, information gathering or control, depending on the specific characteristics of each particular situation" (p. 172).

The notion of the concept is described as a triple consisting simultaneously of a set of *situations* that are meaningful to the concept and where the concept is recognised, a set of *operational invariants* that can be used to deal with these situations and a set of *representations*.

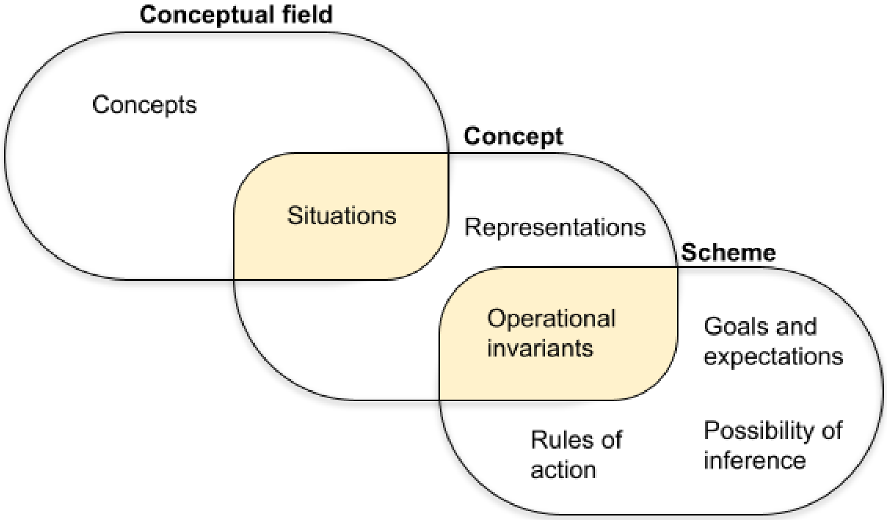


Figure 4: Overview of theoretical constructs related to TCF.

Vergnaud underlined that the meaning of his version of the word 'representation' is similar to a conception of or, at least, a representational reference to a set of situations and invariants. Imagine how thinking about a concept without choosing a relevant representation can seem quite impossible. Vergnaud described how the development of children's schemes depends on the new situations in which they are forced to accommodate their existing schemes, thereby expanding their conceptual field, since the scheme somehow serves as a connection between situations and concepts. In some cases, schemes are not only efficient but also effective, in which case they can be thought of as algorithms. Efficient schemes are not necessarily destined to reach the desired or anticipated goal, where effective schemes (algorithms) are. Vergnaud (2009) defined a scheme as consisting of four aspects or components, as they are often referenced:

> The **intentional aspect** involves a goal or several goals that can be developed in subgoals and anticipations. The **generative aspect** involves rules to generate activity, namely the sequences of actions, information gathering, and controls. The **epistemic aspect** involves operational invariants, namely concepts-in-action and theorems-in-action. Their main function is to pick up and select the relevant information and infer from it goals and rules. The **computational aspect** involves possibilities of inference. They are essential to understand that thinking is made up of an intense activity of computation, even in apparently simple situations; even more in new situations. We need to generate goals, subgoals and rules, also properties and relationships that are not observable.
>
> The main points I needed to stress in this definition are the generative property of schemes, and the fact that they contain conceptual components, without which they would be unable to adapt activity to the variety of cases a subject usually meets. (p. 88, my emphasis on the aspects)

Essential to the schemes are the operational invariants spanning the epistemic aspect of the scheme. A *concept-in-action* is described as a concept that is held to be relevant in the current situation. As a part of every action, we choose certain objects, predicates or categories of such that are perceived to hold relevance in the current situation. A *theorem-in-action* is a proposition held to be true. When we engage in a mathematical situation, we believe certain 'theorems' to be true or false about the objects relevant to the situation. One could identify these theorems as truths derived via mathematical proofs but expanded to truths about objects related to mathematics without. According to Vergnaud, there is a dialectical connection between theorems and concepts;

this emerges from the fact that more advanced mathematical concepts originate from theorems, and vice versa. Nevertheless, it is important to distinguish the cognitive function of the operational invariants in this very specific way. Concepts-in-action are individually available concepts for the enactor-relevant representation in the situation. Concepts-in-action bear no value of logical truth, just relevance. Theorems-in-action are by nature true or false. These entities are sentences (propositions) that provide the concepts with the possibility of inferences taking place. The rules of action are not to be confused with the theorems-in-action. The function of the rules of action is to be appropriate and efficient, but they rely implicitly on theorems-in-action (Vergnaud, 1997).

3.1.3 The schematic process of solving an equation – an example

This section features the theoretical constructs mentioned in the above section applied in an example. The task was presented by Bodin (1993):

$$7x - 3 = 13x + 15.$$

This equation reflects the general form of linear equations mentioned above in section 1.1.1. To engage in the situation and solve the equation, the equation-solving scheme will be activated. The following four aspects will come into play according to where we are in the process. First, goals and anticipations will be set. In this case, the task is to solve the equation, and the goal is naturally adapted from the wording of the task (or the teacher). Second, some rules of action will be set. In this particular situation, the rule of action (strategy) will be to try to isolate the unknown. Of course, the activation of the four components is not as chronological as portrayed in this description. Some concepts-in-action will necessarily be deemed relevant during the process of deciding to isolate the unknown, namely the equation itself (the equals sign), the letter representing the unknown, numbers, operators, etc. Next, we activate theorems fitting the scheme chosen for the situation, specifically the theorems-in-action. To progress, theorems will be applied. To uphold the equality of the equation, the theorem of adding and subtracting equal measures on either side of the equals sign is applied, leaving us with the following:

$$7x - 13x - 3 + 3 = 13x - 13x + 15 + 3.$$

Here, the theorem-in-action give the idea that terms can be rearranged if signs are preserved and that equal terms with opposite signs equal 0 in an expression. In addition, the element 0 can be disregarded if added or subtracted.

$$-6x = 18$$

Next, the theorem of dividing by a non-zero/non-infinity amount on either side of the equals sign is applied. This leads to the following equation:

$$x = \frac{18}{-6}$$

$$x = -3$$

A minor calculation theorem of dividing by a negative number is applied. Now, the unknown is isolated, and the rule of action has been carried out. The inferences will, in this example, be that the equation has a set of solutions that is {-3}.

3.2 Research hypotheses as a means for working with the research questions

In this section, I outline the research hypotheses that informed my exploration of the posed research questions. To do this, I return to the research questions and dedicate a subsection to each, explaining the thought process behind the development of the research hypotheses and the associated didactic variables.

The didactic variables are the micro and macro choices that contribute to the different designs included in the project and the formation of the proposed hypotheses:

> Conception [design] and a priori analysis is a crucial phase of the methodology. It relies on the preliminary analyses carried out, and is the place where research hypotheses are made explicit and engaged in the conception of didactical situations, where theoretical constructs are put to the test. Conception requires a number of choices and these situate at different levels. (Artigue, 2015, p. 6)

The choice in these didactic variables was driven by the preliminary analyses at both the micro and macro levels (Artigue, 2014, 2015); "Among the many variables influencing the possible dynamics of a situation and its learning outcomes, didactical variables are those under the control of the teacher" (Artigue, 2015, p. 6).

In relation to the overall research questions, one might argue that only the interventions created on behalf of the detection and diagnostic processes constitute a didactic situation. However, in line with what I have already mentioned about fundamental situations, I choose to apply this idea here, that I require fundamental situations to create relevant diagnostic situations.

In summary, the three posed research questions were all heavily inspired by the framework for addressing difficulties established by the Maths Counsellor Programme (Jankvist & Niss, 2015). I remind the reader that this framework consists of the idea of *detecting* students experiencing difficulties and *diagnosing* the origin of such difficulties before preparing suitable *interventions* to counteract these diagnosed difficulties. With the adaptation of this idea together with the preliminary analyses, I formed the hypotheses and determined the didactic variables informing the design process.

3.2.1 Research Question 1

Research Question 1 is separated into two parts, *a* and *b*. I chose to leave them as two parts since they both work in the direction of creating means for detecting students experiencing difficulties or, in other words, creating (didactical) situations with diagnostic opportunity. I present again the two parts of the research question.

What design principles are appropriate when structuring and designing tasks capable of detecting students facing mathematics-specific difficulties related to learning the concept of linear equations and equation solving?

a. *What principles are appropriate when exploring one specific difficulty (the equals sign)?*
b. *What principles are appropriate when exploring difficulties related to the more general topic of equation solving found in the literature?*

I wish to address didactic situations as a means for entailing the detection of students experiencing difficulties when solving linear equations. In relation to the first research question, I wish to focus the hypothesis and the effort on the design part of tasks suitable for detecting students experiencing difficulties. Based on the preliminary analyses, and perhaps in particular the institutional analysis, the main access to students is through their inputs and thereby the data collected from MatematikFessor. This choice again stems from the aim to establish a data-driven approach to detecting students experiencing difficulties.

Working towards establishing research hypotheses for Research Question 1, I imagine that extended knowledge, established in the preliminary analyses, on the concept of linear equations and particularly the associated difficulties students experience while working with such, paves the way for analysing or establishing the diagnostic potential of tasks involving solving linear equations.

Returning to the idea of diagnostic value and potential, I imagine that in working with students through the means of the MatematikFessor, the border between what detects students experiencing difficulties and what diagnoses students' difficulties can become blurred. In the publications describing the tests (they are referred to as detections tests) used in the Maths Counsellor Programme, Prof. Niss mentioned that should the tests designed for qualitative use be analysed in a quantitative fashion, new perspectives might appear.

As mentioned, the structure of DE was applied retrospectively. Therefore, I also present the reader with hypotheses, as hypotheses presented themselves throughout the project. Even though some hypotheses might have been explored, new hypotheses emerged.

3.2.1.1 Hypothesis 1.1: In MatematikFessor's database, there are already thousands of tasks about linear equations relevant for lower secondary school students. These tasks have enough diagnostic potential in terms of design to detect students experiencing difficulties.

The above hypothesis was established based on the idea that there exists a vast number of tasks involving solving linear equations, and the sheer volume of data points would suffice as a means for detecting students experiencing difficulties. Additionally, I had the idea that these already existing tasks would have collected data in the form of answers for a long time and would have been somewhat ready for analysis. In collaboration with Christian Hansen, I carried out a study where we attempted to cluster the students utilising unsupervised learning methods based on the thousands of tasks involving solving linear equations already implemented in MatematikFessor. However, as Paper A shows, we did not manage to achieve a fruitful result. Therefore, the idea of establishing task design principles based on the characteristics of students' difficulties emerged. I shall return my exploration of Hypothesis 1.1 in the realisations phase (section 1.1.1) but mention that these findings led me to Hypothesis 1.2.

3.2.1.2 Hypothesis 1.2: Designing a collection of tasks with the aim of uncovering the difficulties found in the literature can strengthen the possibility of achieving a set of fundamental tasks with significant diagnostic potential in detecting students experiencing difficulties.

The tasks that were already implemented in MatematikFessor, the ones that served as the basis for the analysis in Paper A, were all designed using the multiple-choice input option. After examining the options that served as distractors, I found that many of these were not necessarily particularly well chosen, meaning that the diagnostic potential and the diagnostic relevance were not high. The distractors were not chosen based on empirical findings, further subtracting from the diagnostic potential and relevance. In addition, I found that many of these equations were quite similar in structure and potential challenges and additionally not categorised beyond the five levels of difficulty within a collection of tasks that MatematikFessor staff were working with. These levels of difficulty are not important in the explorations made in this project. I simply mention them here to demonstrate that we were struggling to group students based on these already implemented tasks. The problem of choosing a preferred input type was a significant design choice in establishing the didactic situations that should serve as the basis for the data collection. In the second part of Paper B, Prof. Jankvist and I established a set of design principles and an exemplary design of tasks with the potential to assess anticipated or known difficulties when solving equations. Several micro choices were made in the conversion of the collected areas of difficulty to a set of equations suitable for addressing such difficulties at an appropriate conceptual level for lower secondary school. The choice fell on using (controlled) variation theory (Marton, 2015; Watson & Mason, 2006) in an attempt to not only generate a suitable structure of linear equations for lower secondary school but also to have the different types of equations with this structure address as many of the difficulties as possible. The idea was that we could somehow create a collection of tasks suitable for online learning environments and with both collective and individual diagnostic potential enabling the detection of students experiencing difficulties solving linear equations.

3.2.1.3 Hypothesis 1.3: Choosing input fields instead of multiple-choice options when selecting/designing task involving solving linear equations will enable opportunities for discussing diagnostic potential, future design principles and knowledge on proper distractors.

The equations that were designed and implemented based on the findings in Papers A and B were deliberately kept short and simple in formulation, following the idea that the tasks had already been implemented in the environment to avoid loss in diagnostic potential and value by creating an unwanted space for irrelevant difficulties. All the tasks were implemented, resembling the following example:

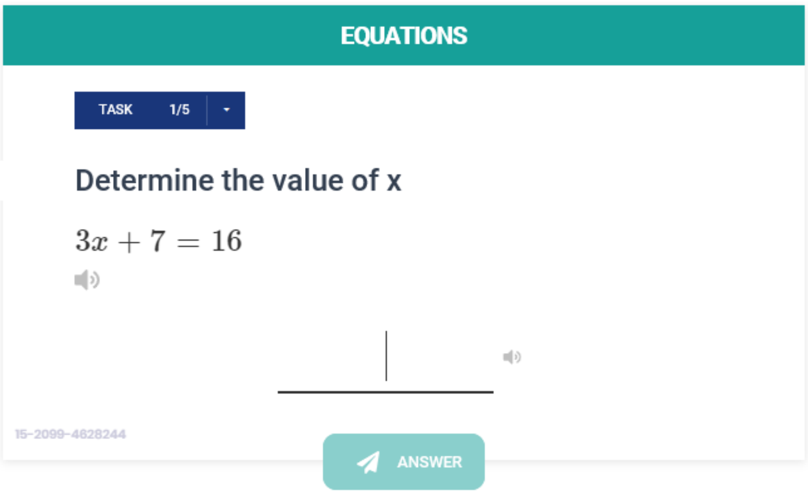


Figure 5: Example task from MatematikFessor [translated from Danish].

Regarding the second part of Research Question 1, I invite the reader to join me in a theoretical design experiment. As part of the educational journey of working as a PhD student, my fellow PhD students at the Danish School of Education and I were introduced to TCF. The theoretical constructs caught my attention because I saw that it could be linked to working with errors and difficulties through the notion of scheme as presented through the TCF. During the PhD course on TCF we hosted at our institution, we acquired extensive insight into this theory. The four aspects (components) of the scheme in particular caught my attention. Much of Vergnaud's work only placed a small emphasis on three of the aspects, since the operational invariants in many cases seemed to be considered the most important or the only noteworthy aspect of the four.

In Paper C, we provide a theoretical framework for designing tasks for implementation in an online learning environment specifically to work with formative assessment and fundamental situations. Since I decided that working with students directly, as prescribed by the framework of the Maths Counsellor Programme, was not a priority, my focus on establishing a framework for working with tasks as a part of diagnostics became a priority. In Paper C, we do not discuss the didactical aspect of the situations in relation to TDS; rather, we focus on the possibilities for generating means for formative assessment (diagnostic assessment) using the ideas from the TCF and discussing knowledge as presented via students' schemes (their actions). The driving hypothesis for the second part of Research Question 1 is as follows:

3.2.1.4 Hypothesis 1.4: The four aspects that define the scheme as a part of TCF can serve as a comprehensive means for designing tasks to ensure diagnostic potential in an assessment situation incorporating an entire action.

I wanted to determine situations that allow for a diagnostic (and didactic) sequence or situations where students and teachers together manage to explore the origin of the difficulties leading to erroneous answers to tasks involving equation solving.

3.2.2 Research Question 2

The second research question has to do with utilising the data from MatematikFessor to potentially assess and analyse students' difficulties in terms of diagnostic value. Again, I chose to explore the research in two parts. I remind the reader of Research Question 2.

What possibilities for the general diagnosis of students' difficulties related to equation solving can be established?

a. *What possibilities can be established based on a task found in the relevant literature?*
b. *What possibilities can be established based on the analysis of an exhaustive set of tasks involving equation solving?*

Following the ideas presented under Research Question 1, my intention is that Research Question 1 is concerned with the design and diagnostic potential of tasks, whereas Research Question 2 is concerned with exploring possibilities for the analysis and diagnostic value of tasks. In the above section, I explained how Paper A in some way serves as a qualifier and a part of the preliminary analyses that, to a large extent, justifies the design of new tasks instead of using already implemented tasks. In relation to Part b of Research Question 2, I propose the following hypotheses based on the idea that with the new tasks I had implemented in MatematikFessor, I would now be able to use a clustering method to categorise students as equation solvers. To some extent, this is an extension of Hypothesis 1.1 but with a different purpose, as it is being viewed in retrospect. Following the lessons learned regarding the task design, I wanted to explore these potential categorisations that could possibly be found in a large data set.

3.2.2.1 Hypothesis 2.1: Extensive didactic coding of answers to tasks can help create a means for exploring the extent of different reasons for students' erroneous answers based on answers to linear equations from MatematikFessor.

Following the design mentioned in Hypothesis 1.2, I aimed to establish what I refer to as reasons for the answers students provided to the tasks based on the design principles established in Paper B. This hypothesis leans into the idea that it is possible to establish the diagnostic value of each of the types of tasks. Remember that the equations were designed in a reasonably exhaustive fashion suited for students in lower secondary school. The coding should reflect these reasons for the errors students make when solving the linear equation, and a part of the hypothesis is that it is possible, to a large extent, to establish these reasons. Furthermore, the idea of going beyond correctness in establishing diagnostic value can help online learning environments provide better and more useful feedback to teachers.

3.2.2.2 Hypothesis 2.2: Unsupervised learning methods (specifically MIRM[3]) together with additional didactic coding of answers to tasks can help create a means for categorising students based on answers to linear equations from MatematikFessor.

Similar to the idea presented under Research Question 1, I assume that a critical number of tasks presented to a critical number of students can result in extensive knowledge of students' actions (schemes) when solving linear equations. One way this idea can benefit the didactic situation is that the knowledge generated based on data could be utilised to support a more detailed diagnostic process than a diagnostic purely based on basic statistics through analysing the correctness of the answers the students provide. An obvious limitation of this hypothesis is that the data consisting of answers to tasks already implemented in the online learning environment were deemed insufficient. The implication of this was to establish additional means for generating data through the design of additional or new tasks to establish grounds for healthy and fruitful data analysis. In line with the previous hypothesis, I wanted to utilise didactical coding that enables the analyses to go beyond correctness. I believe that another strength lies in the implementation of more advanced analytical tools in this coding process that goes beyond correctness. A limitation connected to designing and implementing new tasks is that these tasks require additional time to gather data in the form of answers.

3.2.2.3 Hypothesis 2.3: A diagnostic value can be interpreted by utilising the extensive amounts of data insights provided by applying unsupervised learning methods to data extracted from MatematikFessor.

These questions can address the macro and micro choices in determining the prerequisites for gathering data. Initial macro choices in this context would be the choice of solely utilising data and the target group of the tasks being lower secondary school students in Denmark. In Paper A, Christian Hansen and I concluded that the equations already implemented in MatematikFessor served as insufficient grounds for clustering students into groups based on the answers provided. I must admit that during this phase, I was still establishing the

[3] The Multinomial Infinite Relational Model is a model for co-clustering forming homogenous groups in a data points forming rows and columns. The model is explained in further detail in Paper E.

grounds for the data analysis I hoped to perform. Nonetheless, based on the process of working with the data in Paper A, I decided to conduct a new data analysis with a new collection of equations.

Furthermore, when studying the literature, I realised that tasks of different kinds have been presented in several studies. One particular task and the results presented are particularly interesting to me. The task $8 + 4 = \_\_ + 5$ presented by Falkner et al. (1999) produced very interesting results. The answers the students provided were few and seemed to represent a good idea of how the particular student providing the answer interpreted the task, especially the role of the equals sign, but also other important concepts. This led me to the next hypothesis regarding Part ***a*** of Research Question 2.

3.2.2.4 Hypothesis 2.4: The task $8 + 4 = \_\_ + 5$ (Falkner et al., 1999) holds high diagnostic value and will serve as a good measure for comparing diagnostic value related to understanding the role of the equals sign.

I found during my review of the literature that not many tasks are presented like classic textbook tasks, as many might know them from mathematics textbooks or maybe to a higher extent from digital environments, such as MatematikFessor. In the case of the task presented by Falkner et al. (1999), the empty space represents the unknown instead of a letter. Many might associate the empty space with the lower grades and with tasks highlighting arithmetic procedures and not algebraic (generalised) thinking. It was important to me that the tasks I created would resemble 'normal' everyday tasks that would be recognised as such by students. However, I also wanted to explore how tasks such as $8 + 4 = \_\_ + 5$ would work in a digital version and whether I would get different results from the original authors.

3.2.3 Research Question 3

The third research question is focused on the last step of the Maths Counsellor Programme's framework: the intervention. Research Question 3 was formulated as follows:

What possibilities for general interventions in relation to the concept of linear equations and equation solving can be established?

a. *What possibilities can be established based on a specific difficulty from the literature?*
b. *What possibilities can be established based on difficulties measured in an exhaustive set of tasks involving equation solving?*

I want to discuss the conceptions of didactic situations suitable for interventions against known difficulties related to students' equation-solving schemes. I knew when embarking on this PhD project that I might not be able to base the interventions on the concrete difficulties that the students were diagnosed to be experiencing. Therefore, I chose to concentrate the possible intervention design on two different efforts or paths.

To explore the space of general interventions, some initial questions arose: Is it a good idea for interventions to be generalised or how might such a concept exist to fruitfully target difficulties related to the concept of linear equations and equation solving? To accept the idea of a more general intervention, one might be forced to create or design a space in which it would be possible to address or counteract certain difficulties. This led me to the first intervention designed. As I mentioned in the sections about included papers and collaborations, during my employment at Edulab (the host company), I got the chance to work with Lui A. Thomsen in designing and developing a VR application for teaching and learning linear equation solving with a special focus on working with equations made abstract by negative numbers. Such equations were chosen specifically because this could yield some interesting research aspects for Lui as well.

3.2.3.1 Hypothesis 3.1: It is possible to set up an alternative (VR) environment in which teachers and students can explore equation solving while utilising metaphors to help guide students' equation-solving schemes.

The other path is more in line with the Maths Counsellor Programme, since this framework really aims to build interventions based on the findings from analysing the answers provided to tasks and interviews. However, returning to the issues raised in the introduction, I wanted to generate new value, meaning and possibilities based on the data collected specifically from MatematikFessor. Based on the data, I observed several obvious paths to follow when generating space for counteracting mathematics-specific difficulties related to solving linear equations. I could generate better feedback for students and personalising their learning by applying knowledge about their difficulties in a feedback system. However, as I found in the preliminary analysis, teachers experience difficulties interpreting feedback about students from online learning environments and perhaps struggle to establish formative assessment ideas based on this feedback. The preliminary analyses, especially the institutional analysis, revealed that teachers using the statistical tools in MatematikFessor struggle to make sense of them. Few teachers use the overview to measure performance and base their teaching or evaluations on it.

3.2.3.2 Hypothesis 3.2: Utilising large amounts of data in addition to extensive coding and co-clustering of answers to tasks involving equation solving will pave the way for formative assessment opportunities, allowing teachers to address students' difficulties in solving linear equations using MatematikFessor.

The idea is that multiple analyses of answers in large quantities will generate possibilities for interpretation, building on the entire structure of knowledge collected before to generate reasonable design ideas for implementing better ways to provide feedback for teachers working with online learning environments.

3.3 How the included papers connect to the proposed research hypotheses and didactic variables

In this section, I present an overview of the included papers' relevance to the established research hypotheses. The idea is to provide the reader with summaries of the included papers that are more related to the reading of this *kappa* than the summaries presented earlier in section 1.6.

Paper A presents what came to serve as an initial study or analysis of tasks. In this paper, we present a range of tasks picked from among the already implemented tasks in MatematikFessor. This paper ends up deserving a spot that is not far from the preliminary analyses. The reason for Paper B having two parts is also because of the conclusions made in Paper A. We present a novel attempt, via an unsupervised learning method, to group or cluster students based on their answers provided to tasks implemented in MatematikFessor before the beginning of my PhD project. The paper also provides novel insight into utilising data on students' answers to tasks involving equation solving. In conclusion, the paper catalysed the establishment of new design principles for analysing students' erroneous answers when solving linear equations. Paper A deals mostly with Hypothesis 1.1.

Paper B presents a literature review that led to the design of a collection of linear equations suitable for online learning environments and students in lower secondary school in Denmark. The design includes 11 general types of linear equations—10 types of arithmetical equations and one type of algebraic (non-arithmetical) equation (Filloy & Rojano, 1989; Vlassis, 2002). In addition, we present a broad range of variations of these embedded in a potential learning-trajectory-tree structure covering what I consider a reasonable attempt at a comprehensive set of equations creating fundamental situations for learning linear equation solving. In this way, the literature review, presented in the first part of Paper B, plays a central role in the designs presented in Papers C, D, E and F. When speaking in terms of TCF, the design attempt presented in the second half of Paper B covers the conceptual field of linear equations in relation to lower secondary school students in Denmark. Controlled variation (Watson & Mason, 2006) is utilised in two dimensions—both to secure a reasonable variety of types of equations but also a reasonable variety of equations within each type, addressing as many of the identified difficulties as possible. Paper B plays a huge part in establishing the possibility for further design. It also plays a part in the preliminary analysis part of the DE structure presented in this dissertation. Therefore, Paper B mostly deals with Hypotheses 1.1, 1.2 and 1.3, but it also acts as a prerequisite for most of the project.

Paper C presented an alternative or a more nuanced perspective on task design, featuring tasks that potentially hold a new and different value in terms of diagnostic assessment. The paper presents a framework for designing diagnostic tasks for online learning environments, using TCF and the components of the scheme as a guide or framework. Furthermore, the paper presents a framework for further developing this task design into something relevant for teachers in the classroom. The purpose of the tasks is to enable teachers to better hypothesise about the reason for their learners' errors. In contrast to Paper B, Paper C has a much more direct and qualitative approach to addressing formative assessment and diagnosis through tasks in online learning environments. The paper relates to Hypothesis 1.4.

Paper D replicated the famous $8 + 4 = \_\_ + 5$ (Falkner et al., 1999) task through a modified conceptual replication study (Aguilar, 2020). The paper shows that online learning environments can help create data that help establish new or extended knowledge on how diagnostic tasks perform. The idea was to redesign and reuse elements from highly cited papers, including tasks involving linear equation solving from mathematics education research. The experiment was done to explore another research-supporting feature of online learning environments' capability of gathering data for use in quantitative research. More importantly, perhaps the study helps in relation to the overall project to exemplify diagnostic potential and value. The paper relates to Hypothesis 2.4.

Paper E presents the design of a framework for enhancing data gathered from answers to the collection of linear equations designed and presented in Paper B. Additionally, we present the application of an unsupervised learning method to co-cluster data matrices (students × tasks) to categorise lower secondary school students as equation solvers, as well as the 892 tasks designed based on Paper B. The idea in relation to TCF is to analyse and classify these situations involving linear equation solving and the procedures the students use to deal with them to generate qualitative formative assessments based on difficulties and not on correctness. This paper therefore relates to Hypotheses 2.2, 2.3 and 3.2. Paper E represents work performed at the very end of this PhD project and is therefore not yet a published paper. Establishing possibilities for analysing data, such as the data presented in Paper E, has proven to be quite a time-consuming process. The work presented in Paper E is thus still in the ideas stage, offering some thoughts about improved digital formative assessment at an early stage.

Paper F presents the design and development of a digital experience for teaching linear equations using a modified balance model for equation solving. Lui A. Thomsen and I modified the balance model to alter physics behaviour in a VR experience to strengthen students' schemes for solving linear equations and to help students adapt their schemes to situations where negative numbers and mathematical negativity make equations abstract. The paper should be viewed as an attempt to design and develop a more general possibility for teachers to host interventions for their students, in line with research Hypothesis 3.1. The paper includes an exploratory study aimed at analysing and exploring the effects and affordances of teaching with the modified balance in the VR application. This novel teaching experience was conducted with 10 lower secondary school students.

Chapter 4 Realisation, observation and data collection

In this chapter, I present the realisations, observations and data collection processes for the PhD project as part of the structure outlined by DE. In this phase, I paid attention to whether the data could inform the goals or hypotheses of the a priori analysis. To summarise, Artigue (2015) emphasised the importance of realisations made during this phase, as they often lead to some adaptation of the design, especially when the DE is of a significant size. Because I used DE as more of a structure, I did not include some sections regarding iterative processes in the research design.

In the following section, I attempt to clarify the valuable contributions made by the papers. I also attempt to structure these in order of consequence or the impact the findings have had early in the project's life span. A more detailed discussion on the implications of the findings related to the overall project will be conducted in the a posteriori analysis in part 4. For this part, I focus on the implications of exploring the hypotheses.

4.1 Hypotheses related to Research Question 1

The posed hypotheses presented through these sections are related to Research Question 1, and under each section, I attempt to clarify the results and realisations I made when exploring them. In each section, I remind the reader of the related hypotheses.

4.1.1 Hypothesis 1.1

In MatematikFessor's database, there are already thousands of tasks about linear equations relevant for lower secondary school students. These tasks have enough diagnostic potential in terms of design to detect students experiencing difficulties.

The outcome of exploring this hypothesis has already been touched upon, as I wished to explain some of the thought processes that led to the further hypotheses. However, in this section, I want to dive a little deeper into what we realised when working with the tasks already implemented and some key aspects of analysing them.

As we mentioned in Paper A, regarding the research questions, the aim was to qualify the analyses of tasks and answers to create a method for diagnosing lower secondary school students' difficulties related to equation solving. For that reason, the working question in Paper A involved establishing some categorisation of students based on this analysis, and this categorisation was then to be the subject of further analyses. However, the categorisation, or the clusters, we wanted to establish did not exist based on the knowledge and data we had at our disposal. During the clustering, I took a closer look at the structure and design of the tasks that we had chosen as representing equations suitable for lower secondary school students. The structure can be seen in Figure 6. The equations were what were implemented in MatematikFessor prior to this project and were all using multiple-choice as input method.

Label	Form	Steps	Number of tasks
1	$ax + b = c, \quad x, a, c \in \{1,2, \dots ,9\}, b \in \{0,1, \dots ,9\}$	One/Two-step	106
2	$ax + b = c, \quad x, a, c \in N \; b \in N_0$	One/Two-step	812
3	$ax + b = c, \quad x, c \in Z \; b \in N_0, a \in N$	One/Two-step	501
4	$ax + b = cx + d, \quad a, b, c, d, x \in N$	Two/Three-step	1478
5	$ax + b = cx + d, \quad a, c \in N \; x, b, d \in Z$	Two/Three-step	519

Figure 6: Structure of equations included in Paper A.

The task designers and editors at MatematikFessor are required to set a level of difficulty for each of the tasks implemented in MatematikFessor. This level of difficulty ranges from 1 to 5, where 5 is considered the most difficult. The idea is that the environment can then, based on this difficulty scale, apply algorithms and features in task assignments to students. We found that the level of distinction between the tasks was still too narrow. I realised that I needed to be able to produce means to collect data that were not going to fall into the same set of constraints causing the data to be too similar in construction. In other words, I needed a system or a collection of equations that was more likely to include not only fundamental tasks for solving equations but also equations that could be didactically distinguishable attempting to ensure high diagnostic potential. Therefore, I set up a structure for how I felt I could explore a larger variety of linear equations that could be considered closer to a set of fundamental equations for lower secondary school. I invented the following structure, in the form of a conceptual map, in an attempt to compensate for the similarity in task construction (see Figure 7) based on the findings from the literature review.

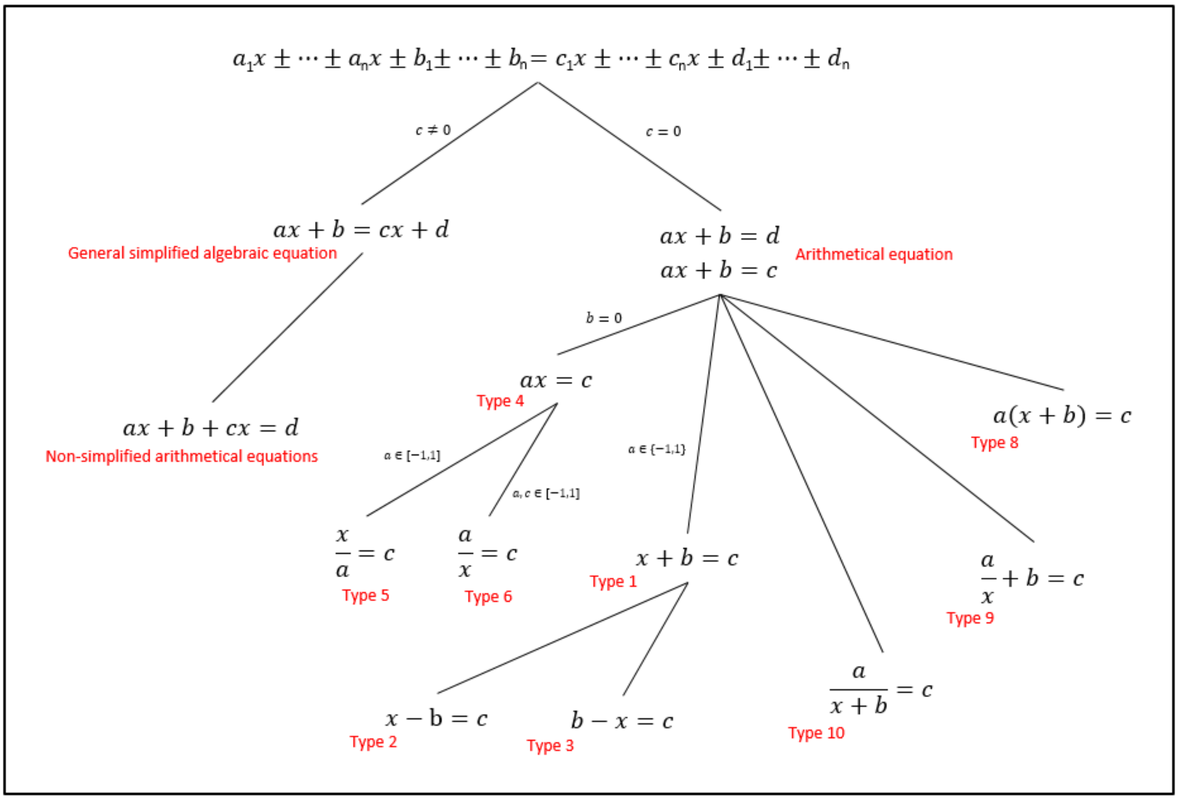


Figure 7: Final conceptual map of linear equations based on the idea of Paper B refined in Paper E.

The conceptual map presented in Figure 7 is the final iteration of the idea presented in Paper E. Paper B presents a slightly different map, where the algebraic equations were not present. In Paper E, I outlined the design, and during the implementation process, I determined what equations we could and should implement.

4.1.2 Hypothesis 1.2

Designing a collection of tasks with the aim of uncovering the difficulties found in the literature can strengthen the possibility of achieving a set of fundamental tasks with significant diagnostic potential in detecting students experiencing difficulties.

From working with the tasks already implemented in MatematikFessor, I realised that to achieve something that could bring me closer to establishing possibilities for providing teachers with new opportunities for feedback, I had to generate means for 'better' data. As I was working with the tasks already implemented, I began to think more and more about how these tasks were designed and what principles went into that process.

Analysing the tasks already implemented in MatematikFessor, I found that what I now call the diagnostic potential was probably too weak when viewed and analysed as a collection. When I say weak, I mean that through talking to colleagues at Edulab, I found that only rather simple design principles had been implemented in relation to the construction of these tasks. Additionally, I found that these tasks were among the first tasks there had ever been designed for the online learning environment back in around 2010. This, alongside the fact that there were a vast amount similar tasks, challenged the clustering we wanted to use to characterise students via their erroneous answers. Another factor, and most importantly, the numbers used for coefficients and other terms in the equations were randomly chosen. This led me to believe that for the diagnostic potential to be significant, the numbers would have had to been chosen in a rather lucky manner for any given task to actually capture problematic characteristics. An example could be to showcase the reasoning behind two rather similar tasks with quite different diagnostic potentials, in my opinion Kieran (1985) provided some of the most down-to-earth explanations of what students actually do when making errors solving linear equations. From her ideas/findings, one could extract important design principles. These are subtle but important when designing an extensive number of tasks. I realised the potentially important, perhaps subtle, differences allowing for inversing subtraction with subtraction, when solving equations, per design. The following example showcases this small but important realisation. One huge disclaimer is that task designers do not really know how a task performs and whether some unexpected behaviours in the form of strange answers might present themselves until the answers are collected and analysed.

1. $5x - 6 = 14$

In this first example, if a student were to inverse the subtraction by subtraction, the action might be regretted simply because the next expression does not make good sense if the classroom is not used to answers such as 8/5. I know that additional errors might cause a student to give the answer 3 simply because 5 plus 3 makes 8.

2. $2x - 7 = 21$

In the first example, though an interesting task, inversing the subtraction with subtraction is not the focus in the same way as it is in this second example. This equation makes sense (for positive whole number solutions) whether subtraction is inversed with subtraction or not. This led to the idea that the ongoing literature review should actually be leading to a set of design principles, hence the second part of Paper B. This realisation I consider quite important for the further design choices I made. Working with the review of students' difficulties, I established the following design principles for creating new tasks (Paper B):

- Negative numbers and the minus sign
    - as solutions
    - as terms
    - as operations
- Rational numbers
    - as solutions
    - as present numbers
- Interpretation of the equals sign
    - Situations that invoke an 'element of crises'[4] from an arithmetic point of view
- Strategic, conventional and transformational questions
    - Increasingly complex and strategically demanding
    - Conventions concerning brackets
    - Conventions concerning missing multiplication

Following these design goals (principles), I attempted to realise these in a concrete design for each type of equation presented in the conceptual map (Figure 7). This paper presents concrete examples of how such design goals could be manifested in linear equations suitable for implementation in an online learning environment. However, we needed a structure for this manifestation not provided by the design principles. The structure was designed using an exhaustive method partly based on variation theory. The structure was presented prior to the final implementation of the equations in the online learning environment. Working with the team of professionals in the editorial team at Edulab, we ended up not considering Type 8 (equations with brackets), as we constructed and implemented the equations into the online learning environment.

4.1.3 Hypothesis 1.3

Choosing input fields instead of multiple-choice options when selecting/designing task involving solving linear equations will enable opportunities for discussing diagnostic potential, future design principles and knowledge on proper distractors.

Continuing with the idea of diagnostic potential, when realising that the tasks already implemented in MatematikFessor were probably not going to lead to a fruitful investigation relevant to the overall aim of exploring new data-driven possibilities for feedback under the umbrella of the Math Counsellor Programme's framework of detecting, diagnosing and intervening, I began to think more and more about task design and its importance in establishing means for diagnostics in online learning environments.

When I began establishing design principles for new and preferably more suitable tasks to realise the overall idea of enhancing feedback offered by online learning environments, I found myself taking a few side steps from the data analysis track I thought I was going to follow. It was not until much later in this project that I was able to establish a formulation or definition of what I consider diagnostic value or potential. However, the fact that the clustering of the tasks already implemented did not yield a fruitful result caused me to wonder how a good task should be presented. In MatematikFessor, tasks must meet certain requirements, such as anticipated difficulty level (from 1–5), and every task should be connected to an explanation that students would be presented should they give a wrong answer. Following the guidelines and structure of tasks presented by MatematikFessor, I began to design new tasks based on my initial findings from the literature review. I realised through this process that the multiple-choice possibilities had never been analytically validated in terms of determining what I now consider the diagnostic value of the task. These multiple-choice options were

[4] The element of crisis in task design aims to contrast with standard procedures or understanding students usually apply (Bokhove, 2017).

simply based on a few people's experiences and assumptions on what answer students might be willing to choose.

Through conversations with Prof. Hodgen and Prof. Küchemann, I realised that to truly determine the best multiple-choice options or distracters, as some would refer to them, one would necessarily have to determine these through empirical testing. The two professors argued that to diagnostically test a student purely based on them solving tasks, one would have to present to them a too-large number of tasks for a test. However, one could counteract this by presenting the students with 'perfect tasks' precisely adjusted to measure certain parameters because of their unreasonably high diagnostic value. 'Perfect tasks' would only receive two different answers the correct answer and an incorrect answer perfectly exemplifying a recognised mathematical reason or action. I believe it is fair to assume that such tasks do not exist. Therefore, tasks designed for diagnostic purposes would have to accomplish the fine act of balancing both quantity and quality to achieve success in performing diagnostics.

When working to detect and diagnose students' difficulties in general terms, I realised that there is sometimes only a fine line between detection and diagnosis, knowing that detection refers to the idea that there is some unsuccessful handling of a situation and diagnosis refers to the idea of knowing why this handling was unsuccessful. I realised that when designing tasks for either purpose, I found myself in a place of caring about both. I found that the idea of an unfocused task for detecting students experiencing difficulties could generally just be any linear equation. Then, I would argue that the possibility of accidentally not detecting students who one would actually want to detect would be too high. I wanted to somehow make sure that I formulated types of equations sufficient for providing students with enough different situations to cover the topic of equation solving and for them to be able to actually experience difficulties in every situation. This might boil down to presenting students with every linear equation type and variation possible. However, this could be too comprehensive to accomplish. I argue that this is exactly the idea that ties detection and diagnosis so tightly together. When I create enough situations for teachers or observers to detect students experiencing difficulties, I would have implicitly created means for some level of (a priori) diagnosing to take place. This argument I believe continues to the next level of design. If one could design enough situations for students to display their difficulties, then one would also realise that within these situations, there should necessarily also be opportunities to display certain characteristics.

In choosing the input field for the tasks I designed based on the literature reviewed and choosing some appropriate design principles that would fit the framework for how tasks would be presented in the online learning environment, I realised that I would be faced with the challenge of managing more answers than I would have to using multiple choice. However, adhering to the ideas of diagnostic value, diagnostic potential and through discussions I had with Prof. Küchemann and Prof. Hodgen, I believe that this decision allowed me to achieve a more meaningful analysis of how the tasks perform in situations for diagnostic purposes. I return to the importance of this when discussing Hypothesis 2.1.

4.1.4 Hypothesis 1.4

The four aspects that define the scheme as a part of the TCF can serve as a comprehensive means for designing tasks to ensure diagnostic potential in an assessment situation incorporating an entire action.

In following the path of determining the task design philosophy and principles, I went on to think about a method for discussing perceptions and knowledge of subjects undergoing diagnostic evaluation. Up until I attended the PhD course about TCF, I felt that I had not found my epistemological standpoint when thinking about (mathematical) knowledge. As I mentioned in the preliminary analyses, the matter of discussing the origin of difficulties related to mathematics is not straightforward. I mentioned TDS and TCF as my most prominent sources of inspiration when discussing and analysing knowledge and didactical design in mathematics education.

I was fascinated with the idea that knowledge needed to be applied for an instrument to measure it. In my endeavour to detect and diagnose students' difficulties solving equations, I adopted the idea of analysing students' schemes as a way of capturing their actions that lead to application of mathematical knowledge when solving linear equations unsuccessfully. When discussing the notion of scheme in terms of Vergnaud, I realised that in his terms, an action (a scheme) consists of four elements. I felt that I had found a theoretical standpoint and a new outset for describing knowledge.

During the PhD course on TCF, I discussed the notion of scheme with experts on the theory. I discovered that apparently, there is not the same interest in nor emphasis on the importance of three aspects of the scheme (the intentional, generative and computational) compared to the operational invariants (the epistemic aspect).

Although Paper C is rather reflective of a thought experiment in writing, the idea of gathering information about an action and possibly a specific part of an action to diagnose students' difficulties with mathematics is what resulted from going through the process of making this framework for designing tasks with Prof. Hodgen. The idea that the cause or the origin of a student's error could stem from different aspects of the action (scheme) and that TCF gave us the means to discuss and think about this is what drove the establishment of the framework for implementing the a set of general design principles for diagnostic tasks for online learning environments.

4.2 Hypotheses related to Research Question 2

In the following sections, I present the realisations that came from exploring the hypotheses related to Research Question 2. In each section, I remind the reader of the related hypotheses.

4.2.1 Hypothesis 2.1

Extensive didactic coding of answers to tasks can help create a means for exploring the extent of different reasons for students' erroneous answers based on answers to linear equations from MatematikFessor.

When exploring the potential of the new tasks presented in the design proposed in Paper B and analysed in Paper E, I realised a variety of different things. To explain these realisations fully, I first want to address some initial problems that had to be dealt with to explore this hypothesis.

I anticipated that one obvious obstacle resulting from the extant feedback system is that the system does not offer any information as to why any given task is answered incorrectly. I wanted to explore the possibilities for performing statistical analyses on data sets consisting of students' answers to tasks that went beyond correctness. The idea that online learning environments can process data before presenting the users with automated feedback is unique to these digital resources, meaning they have powerful potential, in my opinion.

The data set I ended up with from releasing the equations I designed into MatematikFessor consisted of 2,135,968 unique answers to a total of 892 unique tasks. I had removed 373,384 answers from the raw data set to end up with a situation where each student had answered each task a maximum of once. I decided that the most recent answer was the one I would keep in the set. The answers were provided by 94,368 students. There is value to be gleaned from the duplicate answers. Students' progression and patterns answering the same task could be exciting to analyse; however, it was out of the scope of this project. The data set was then coded for the five most popular answers to each of the 892 tasks using the following coding format whose construction began prior to the coding but was developed during the coding as well. The coding resulted in 49 unique codes constructed from the base codes in Table 2.

Table 2: Base set of codes with explanations as presented in Paper E

Name	Tag	Explanation
Correct	c	Correct answer
Negativity issue	n	Answers with additive inverses
One more/one less	o	Answering for example 6 instead of 5
Multiplication	m	Issues involving multiplication
Addition	a	Issues involving addition
Rearranging for sense making	r	Rearranging an expression to have it make sense
Equals sign	e	Issues related to the role of the equals sign
Conventions associated with letters	v	Issues related to the interpretation of letters
Solving strategy	s	Issues related to equation-solving strategy
Decimal number	d	Issues related to the acceptance of decimal numbers
Zero	z	Issues related to the number 0
Unknown reason	u	Unable to apply a valuable code

Table 3 shows the most frequent codes. I realised that when I was done coding all the five most popular answers to the 892 tasks that some codes were related to quite a few answers and might therefore seem irrelevant. However, I decided to proceed with the data analysis using the co-clustering method as described in Paper E, since the alternative would show the few occurrences of more generic codes, such as ‘unknown reason’, or eliminate tasks from the data set to streamline the number of codes. I decided to attempt to keep the data as ‘real’ or representative of what actually happened as possible.

Table 3: Distribution of the most frequent codes applied to more than 1% of interpretable errors to the answers to the 892 included equations

Name	Tag	Total answers	% of total answers	% of interpretable errors
Correct	c	1722740	80.7	---
Unknown reason	u	182450	8.5	---
Negativity issue	n	73373	3.4	31.8
Rearranging for sense making	o	36046	1.7	15.6
One more/one less	r	35814	1.7	15.5
Rearranging and negativity issue	rn	27863	1.3	12.1
Rearranging and multiplication issue	rm	13417	0.6	5.8
	av	9098	0.4	3.9
Ignoring the coefficient	mv	6279	0.3	2.7
Rearranging and addition issue	ra	5709	0.3	2.5
Equals sign	e	5511	0.3	2.4
Rearranging and negativity issue with adding the coefficient to the unknown	avrn	3505	0.2	1.5
Disregarding operations on the unknown	v	2833	0.1	1.2

I realised when I extracted the data consisting of the answers to the 892 equations that the answers followed a certain pattern of a few popular answers followed by a tail of less and less popular answers. Because the tasks used an input field, the tail of more and more uninterpretable answers would be unavoidable. I decided that I would prioritise coding the five most popular answers for each of the 892 tasks to provide reasons for each of those answers. In Paper E, some of the realisations connected to exploring the system of coding are presented; however, much is still left for future research. I shall discuss this matter in the a posteriori analysis in Chapter 5 .

By using the co-clustering method together with the didactical codes in writing Paper E, I realised some interesting behaviours that I did not expect, based on reasons derived from research literature. The coded 'o' (accidentally pressing the wrong button on the keyboard when inputting the answer) is something I believe to be unique to digital learning resources. Importantly, if I had not chosen an input field, I would not have been able to observe the phenomenon of the 'one more/one less' issue that presented itself in Paper E. Analysing and coding the answers from the approximately 94,000 students, I found that a significant number of errors to the 892 equations I designed were due to students answering with an unlucky press of a number on the keyboard. In some cases, students must had made reasonable progress in determining that the solution to, for example, the equation $6 - x = 3$ should be 3 but ended up inputting 2 or 4 as the solution. In the following, I present the most frequent error types for each type of equation and present the realisations I made from the analysis. Additionally, I present the common error types that constituted more than 1% of the interpretable errors. I chose not to present the types of equations in numerical order, as I found that the current arrangement was more appropriate for demonstrating the types of errors.

Table 4: Type 1 equations with codes that were assigned to more than 1% of the interpretable answers

Type 1	Total answers	% of total	% of identified errors	Examples
c	564155	88.66	—	$3 + x = 8$
u	33788	5.31	—	$38 + x = 83$
n	15094	2.37	39.37	
o	12669	1.99	33.04	$5 = 1 + x$
r	4764	0.75	12.42	$7 + x = 4$
rn	2635	0.41	6.87	
e	2152	0.34	5.61	
v	758	0.12	1.98	

Here, we see that the identifiable wrong answers that scored high were 'n' (answering with the additive inverse) and 'o' (accidentally pressing the neighbouring button on the keyboard when answering). The reason why we can observe 'e' (issues related to the equality symbol) in this table is because of the task originally presented by Falkner et al. (1999) and the variations presented in Paper D. I found that 'r' (rearranging expressions for sense making) are common in equations where the terms invite individuals to accidentally interpret the expression in a way that is perhaps more familiar or something students are more used to working with. Importantly, I also found that most of the errors to these types of equations were actually uninterpretable.

Table 5: Type 2 equations with codes that were assigned to more than 1% of the interpretable answers

Type 2	Total answers	% of total	% of identified errors	Examples
c	222435	80.90	—	$x - 3 = 5$
u	20396	7.42	—	$x - 9 = -4$
r	10471	3.81	32.60	$x - 4 = 8$
n	9966	3.62	31.03	
rn	7888	2.87	24.56	
o	3517	1.28	10.95	

For these second types of equations, type ‘r’ errors represented a third of interpretable error alongside type ‘n’. Perhaps this is not surprising, since the structure resembles Type 1 equations. The ‘rn’ error (answering with the additive inverse after rearranging) was present with this type of equation, since these equations allow students to inverse subtraction with subtraction (Kieran, 1985). For these types of questions, the diagnostic value would seem quite high, since the results yielded only four quite strong contributions in the form of interpretable error types. I do remind the reader that close to 7.5% of the errors were uninterpretable.

Table 6: Type 3 equations with codes that were assigned to more than 1% of the interpretable answers

Type 3	Total answers	% of total	% of identified errors	Examples
c	193248	79.31	—	$6 - x = 3$
r	18408	7.56	48.52	$3 - x = 6$
n	13553	5.56	35.72	$3 - x = 0$
u	12461	5.11	—	
rn	3514	1.44	9.26	
o	1390	0.57	3.66	
mv	1036	0.43	2.73	

The third type of tasks again featured only three terms with a minus operation present. However, now the unknown is subtracted from a known number. I interpreted the change in the distribution of answers as a sign that these types of equations have a higher diagnostic value than Type 2 equations. The fact that the uninterpretable errors present significantly fewer observations is an important observation. The proportion of correct answers was practically the same as the type before, but the type ‘r’ errors over doubled in the frequency of occurrence. I suspect that Danish students are not so familiar with these types of equations.

Table 7: Type 4 equations with codes that were assigned to more than 1% of the interpretable answers

Type 4	Total answers	% of total	% of identified errors	Examples
c	221265	83.72	—	$2x = 14$
u	19410	7.34	—	$4x = 4$
n	8493	3.21	35.95	$6x = -18$
o	6795	2.57	28.76	$-3x = 15$
av	6319	2.39	26.75	
no	732	0.28	3.10	
rm	706	0.27	2.99	

The fourth type of equations allows the study of equations with two terms and the unknown concatenated with a scalar or a coefficient. Again, the type ‘n’ and ‘o’ errors scored the highest. However, the type ‘n’ error was triggered by the presence of the minus sign and was only an issue in these equations for these types. The error type ‘av’ (confusing concatenation with addition) was present due to the fact that coefficients were introduced and accounted for more than 25% of interpretable errors. This can perhaps be expected, but nonetheless, it is a significant indication that certain error types are connected to the equations based on structure or design.

Table 8: Type 5 equations with codes that were assigned to more than 1% of the interpretable answers

Type 5	Total answers	% of total	% of identified errors	Examples
c	58189	73.72	—	$\frac{x}{3} = 9$ $\frac{x}{-3} = -3$ $\frac{x}{7} = 7$ $\frac{x}{3} = -15$
rm	7332	9.29	51.97	
u	6635	8.41	—	
n	2578	3.27	18.27	
v	1780	2.26	12.62	
ra	1360	1.72	9.64	
rnm	343	0.43	2.43	
rna	177	0.22	1.25	
rmo	168	0.21	1.19	

Type 5 tasks are, to a large extent, the reverse structure of the types before. However, I consider the added feature of the operator (fraction bar) an important difference. The error type ‘rm’ (rearranging with multiplication) was a significant error associated with these types of equations. Importantly, these errors were more common than the uninterpretable errors together. The errors were the result of students simply reversing division with division, using Kieran’s terminology. The proportion of correct answers was significantly lower than in the previous types of equations. The type ‘n’ errors were not surprisingly present here in relation to equations with negative numbers. The type ‘v’ errors (disregarding operations on the unknown) played a bigger role than I expected. These errors were present in response to equations where the denominator and the number on the right side of the equals sign were the same. The error might also rely on the idea that students imply that $\frac{a}{a} = a$. This would, of course, require further exploration to verify. Type ‘ra’ errors (rearranging with addition) comes into play here in situations where students think that adding the numerator and denominator should result in the number on the right side. Notably, the error type ‘o’ did not appear among the high scorers for errors associated with this type of equation.

In Table 9 and Table 10, type 7 equations are presented in two groups, $ax + b = c$ and $ax - b = c$ (where b is always positive). The equations were implemented in MatematikFessor as two types and I found it important to notice the differences in answers the two types would receive.

Table 9: Arithmetical equations (Type 7, $ax + b = c$) with codes that were assigned to more than 1% of the interpretable answers

Arithmetical (+)	Total answers	% of total	% of identified errors	Examples
c	109337	76.04	—	$2x + 7 = 21$
u	17631	12.26	—	$7x + 28 = 7$
o	5054	3.51	30.04	$6x + 7 = -14$
n	4658	3.24	27.69	$-7x + 9 = 30$
mv	2998	2.08	17.82	
av	1631	1.13	9.70	
mvr	713	0.50	4.24	
no	527	0.37	3.13	
rn	463	0.32	2.75	

The proportion of correct answers, for the first type 7 equations, was still fairly high, yet uninterpretable answers rose to above 12%. To my surprise, I found 'o' errors were most common, though unsurprisingly, this was followed by 'n' errors. For the first time, 'mv' (disregarding the coefficient or treating the entire term containing the unknown as the thing we are looking for) became a prevalent type of error. I can best showcase the 'mv' error through the following equation: $5x + 6 = 21$. Students seemingly disregarded the coefficient and gave the answer 15. The same issue occurred with 'mvr' errors, where the student would just rearrange the expression before applying the same logic leading to the same error. The error 'av' was unsurprisingly still present to some extent, since the equations do feature coefficients. The error type 'no' combines the 'n' and 'o' errors and reflects the idea that the students chose the additive inverse as the solution and pressed one more or one less on the keyboard.

Table 10: Arithmetical equations (Type 7, $ax - b = c$) with codes that were assigned to more than 1% of the interpretable answers

Arithmetical (-)	Total answers	% of total	% of identified errors	Examples
c	53395	79.99	—	$2x - 7 = 11$
u	6270	9.39	—	$4x - 4 = 24$
rn	2411	3.61	34.00	$-8x - 13 = 7$
n	1331	1.99	18.77	$4x - 8 = -24$
o	1304	1.95	18.39	$5x - 25 = 5$
mv	544	0.81	7.67	$2x - 7 = 12$
d	279	0.42	3.93	$8 = 2x - 4$
mvrn	242	0.36	3.41	
avrn	207	0.31	2.92	
e	141	0.21	1.99	
av	131	0.20	1.85	
rmv	106	0.16	1.49	
avn	92	0.14	1.30	
mvn	89	0.13	1.26	

For these second type 7 equations, I found a different distribution of answers compared to first type 7. Several were quite common (‘rn’, ‘n’ and ‘o’) before I observed a tail of error types representing the last of the interpretable errors. The presence of ‘rn’, ‘n’ and ‘o’ errors did not surprise me. The ‘mv’ error did surprise me, as it appeared with this equation type as it did for the previous equation type. The versions explain the tail of error types. ‘mv’, ‘mvrn’, ‘rmv’ and ‘mvn’ errors are all connected. The same is true for ‘av’, ‘avn’ and ‘avrn’ errors. The main issues can perhaps be viewed as ‘mv’ or ‘av’ issues, and before the occurrence of these errors, the student had made an error based on rearranging and/or related to negativity. Lastly, I want to touch on the error types ‘d’ and ‘e’. The ‘d’ errors are connected to working with decimal numbers. For example, students would choose to give 9 as an answer instead of 9.5 to the equation $2x - 7 = 12$. This error is also perhaps tied to the idea that if there are no decimal numbers present in the equation, there should not be a decimal number in the solution. Regarding error type ‘e’, I found that some students were willing to put 4 as a solution to $8 = 2x - 4$, indicating that the minus operator and the term ‘4’ to the right of the unknown and the coefficient were disregarded because they operated under the idea that only one ‘thing’ should be present on the right side of the equals sign.

Table 11: Simplified algebraic equations (Type 11, $ax \pm b = cx \pm d$) with codes that were assigned to more than 1% of the interpretable answers

Algebraic	Total answers	% of total	% of identified errors	Examples
c	130189	68.00	—	$7x + 9 = 2x + 4$ $x - 12 = 3x - 2$ $2x - 8 = 7x + 7$
u	34384	17.96	—	
n	10329	5.40	38.43	
rn	8772	4.58	32.63	
avrn	1910	1.00	7.11	
o	1476	0.77	5.49	
mvrn	1297	0.68	4.83	
d	803	0.42	2.99	
mv	616	0.32	2.29	
no	371	0.19	1.38	

Type 11 equations are what I generally refer to as algebraic equations, with the unknown present on both sides of the equals sign. To solve these equations, it is more difficult to apply some substitution techniques or a guess-and-check strategy. Two major error types associated with this question type are highlighted below, as well as what seem to be quite a few uninterpretable answers. Perhaps one would expect the variations in answers to arise, keeping in mind that I only coded the five most popular answers for every equation. However, I think the diagnostic potential of algebraic equations is still valuable. These equations helped me evaluate whether students are capable of handling the unknown’s presence in more than one term and on both sides of the equals sign. Given the design of the equations, ensuring that the variations would cover as many of the found difficulties as possible, I did not find it strange that errors involving improper interpretation and the handling of negative numbers and negativity were among the most prevalent again. Interestingly, the errors connected to treating the coefficient as not existing (‘mv’ errors) or treating it as added onto the unknown (‘av’ errors) mostly occurred together with a rearranging coinciding with a negativity issue error. Additionally, I observed that the error type ‘o’ was still among the most frequent error types to occur, corresponding to more than 5% of identified errors. These types of equations were the first to have a proportion of correct answers below 70%. These equations proved to be the third most difficult among the types I implemented.

Table 12: Type 6 equations with codes that were assigned to more than 1% of the interpretable answers

Type 6	Total answers	% of total	% of identified errors	Examples
c	64768	77.25	—	$\frac{9}{x} = 3$ $\frac{2}{x} = 2$ $\frac{18}{x} = -9$
u	6992	8.34	—	
n	4447	5.30	36.82	
rm	2709	3.23	22.43	
ra	1669	1.99	13.82	
e	1267	1.51	10.49	
zm	606	0.72	5.02	
o	595	0.71	4.93	
ne	262	0.31	2.17	
no	233	0.28	1.93	
m	185	0.22	1.53	

Type 6 equations are inverse in structure from Type 5 equations. Thus, I expected ‘rm’ errors (rearranging with multiplication) to be among the most prevalent. Similarly, the number of uninterpretable erroneous answers was quite low compared to other types of equations, which signifies better opportunities for high diagnostic value. Starting from the bottom, the ‘m’ (multiplication) error was, in every case, connected to the idea that a decimal number would be the solution to an equation such as $\frac{3.3}{x} = 1.1$. For example, students chose 3.3 as the solution. Additionally, I observed the ‘zm’ error (multiplication/division with zero), which refers to students choosing the number zero as the solution to these equations, although only in instances in which equations were in the form $\frac{a}{x} = a$, indicating the idea that a number divided by zero should result in the number itself. Code ‘e’ errors (issues with interpretations of the equals sign) were found for this type of equation where the answers were equal to the number on the right side of the equals sign. In some cases, whether equations were in the form $\frac{a}{x} = a$ or $\frac{b}{x} = a$, some students would choose a as the solution. This issue was observed surprisingly many times and was not limited to the idea that a number divided by itself is the number (for example $\frac{x}{x} = x$). The error ‘ne’ corresponds to errors that involved the additive inverse of the error type ‘e’ that I just discussed. Lastly, I want to mention the errors coded as ‘ra’ (rearranging with addition). This code was applied to instances where students would treat the fraction not as indicating the need for the process of division but addition or subtraction. This error was observed in almost 14% of interpretable errors and is important in the idea of establishing diagnostic value for these types of equations.

To my regret, no variations of the form $\frac{a}{x} = b$, where *a* is larger than *b*, were implemented in MatematikFessor as part of this study. My assumption is that these variations would have held high diagnostic value.

Table 13: Type 9 equations with codes that were assigned to more than 1% of the interpretable answers

Type 9	Total answers	% of total	% of identified errors	Examples
c	52629	72.58	—	$\frac{3}{x}+2=5$ $\frac{8.5}{x}-5.5=-4.5$
u	10096	13.92	—	
rm	2225	3.07	22.73	
e	1796	2.48	18.35	
n	1173	1.62	11.98	
mv	1036	1.43	10.58	
o	1011	1.39	10.33	
avrn	648	0.89	6.62	
rn	504	0.70	5.15	
av	434	0.60	4.43	
d	176	0.24	1.80	
mvn	173	0.24	1.77	
avn	169	0.23	1.73	
rmn	135	0.19	1.38	

Type 9 equations, which are an extension of type 6 equations, yielded only a slightly lower proportion of correct answers. However, these equations yielded a significantly higher proportion of uninterpretable errors. The most prevalent interpretable errors were part of a list where five error types constituted more than 10% of errors each. Again, the more complex in structure the equations are, the more types of errors one might expect. With that said, I saw that 'rm' errors (rearranging with multiplication) constituted the highest number of interpretable errors. Type 'n' errors were again represented, but type 'e' errors occurred more frequently this time. Type 'e' errors was were students would simply give the number on the right side of the equals sign as the solution to the equations. Perhaps not completely related to issues with the equals sign, however quite related to knowing what an equation is and what is meant by a solution to it.

Type 'mv' errors were quite interesting in these equations. For these equations there was a commonality of seemingly interpreting the entire fraction as the unknown, as if the students would carry out a 'cover up' strategy and stopped halfway giving the covered as the solution. These errors are exemplified through the following example. A student gave 3 as the solution to the equation $\frac{3}{x}+2=5$. Using the same idea as the previous type, we can interpret this as the student disregarding the division operation. Whether this idea was applied before or after 2 was subtracted from 5, we will perhaps never know. Unfortunately, in this example, students' could have applied the idea mentioned under a previous type that $\frac{a}{a}=a$. However, In the example of $\frac{8.5}{x}-5.5=-4.5$, I observed behaviour indicating that students would treat the entire fraction as the unknown, leaving the answer 1. I return to the discussion of the diagnostic value of the task $\frac{3}{x}+2=5$ in the a posteriori analysis (Chapter 5). The remainder of the error types were not surprising and had appeared before. Error type 'd' (decimal numbers) appeared with this type of equations as well. I realise that the errors appear in correspondence with the design, but I found it very interesting that students answered 2 or 3 instead of 2.5. Similarly, in other situations, students answered 1 or 2 to a task where 1.5 was the solution.

Table 14: Type 10 equations with codes that were assigned to more than 1% of the interpretable answers

Type 10	Total answers	% of total	% of identified errors	Examples
c	40564	66.98	—	$\frac{4}{x+1}=2$ $\frac{12}{x-7}=-4$
u	11099	18.33	—	
ra	2647	4.37	29.74	
r	1885	3.11	21.18	
n	1746	2.88	19.62	
o	1094	1.81	12.29	
rn	664	1.10	7.46	
rm	389	0.64	4.37	
z	289	0.48	3.25	
d	174	0.29	1.96	

Type 10 equations were again a progression from the $\frac{a}{x}=b$ (type 6) form of equations. These equations proved to be the second most difficult among the ones I designed. Perhaps expectedly, and however unfortunately, the uninterpretable errors summed up to over 18% of all answers. Interestingly, I observed that 'ra' errors (rearranging together with an addition/subtraction issue) were the most prevalent. The error was present in situations where students would resolve or interpret the fraction bar or the fraction construction as either subtraction or addition. In these cases, the students would actually treat the denominator correctly. In the example of $\frac{8}{x+1}=4$, they chose 3 as the solution. The error type 'r' occurred with these equations, signifying that the students rearranged and ignored the constant term added onto the unknown and treated the entire denominator as the unknown. Again, 'n' and 'o' errors were among the most common, with 'rn' errors occurring in cases where students made a calculation mistake with the denominator by attempting to reach the correct value as a dividend. For example, for $\frac{-20}{x-9}=4$, students would not choose 4 but 14 in the attempt to reach 5 or -5. Interestingly the error type 'z' (issues related to the number zero) showed up for these equations. I observed issues arising from when students attempted to divide by 0 in response to equations in the form $\frac{a}{x+b}=a$. The students chose a number that would make $x+b$ equal zero as a solution. My interpretation of this phenomenon is that some Danish students believe that a number divided by 0 is the number itself.

Table 15: Non-simplified arithmetical equations (Type 12, $ax \pm b \pm cx = d$) with codes that were assigned to more than 1% of the interpretable answers

Non-simplified arithmetical equations	Total answers	% of total	% of identified errors	Examples
c	12304	66.62	—	$x+2+2x=8$ $4x-11-2x=-3$
u	3184	17.24	—	
rn	983	5.32	32.96	
o	971	5.26	32.56	
avrn	495	2.68	16.60	
av	347	1.88	11.64	
v	101	0.55	3.39	
mv	46	0.25	1.54	
mvrn	34	0.18	1.14	

Type 12 was the last type of equation I designed surprisingly they proved to be the most difficult for students to solve. I refer to these as non-simplified arithmetical equations. Unfortunately, the proportion of uninterpretable errors was also quite large for this subset of answers. However, with these equations, I found that the error of choosing the additive inverse as the solution (type ‘n’ errors) did not constitute more than 1% of interpretable errors. The two most prevalent error types are common errors, specifically ‘rn’ and ‘o’. I was surprised to find that type ‘o’ errors constituted such a huge proportion of the interpretable errors. Perhaps not surprisingly, the error type ’rn’ was significantly prevalent among the responses to equations with many terms and with occasional operational minus signs. Type ‘av’ and ‘mv’ errors were also unsurprisingly present; however, ‘v’ errors (issues related to the handling of the unknown) did come up in examples where students would seemingly disregard both the terms containing the unknown and provide a solution that would, together with the term not containing the unknown, sum up to the number on the right side of the equals sign.

4.2.2 Hypothesis 2.2

Unsupervised learning methods (specifically MIRM) together with additional didactic coding of answers to tasks can help create a means for categorising students based on answers to linear equations from MatematikFessor.

Hypothesis 2.2 was based on the idea that even though I was unsuccessful in clustering students based on the linear equations already implemented in MatematikFessor, I still thought that the clustering idea could work. I wanted to create the best possible conditions for this idea to work. Together with the staff at MatematikFessor and my company supervisor Klaus Pedersen, I implemented the equations formulated based on the design principles presented in Paper B.

I soon realised that this data set was extremely sparse, since the number of users who had interacted with one or more tasks was 94,368. To illustrate how sparse the data was, we looked at the (student × task) matrix and found that this matrix had 84,176,256 entries. This meant that I only had about 2.5% of the data that would complete the student × task matrix. As explained in Paper E, this matrix contained all the corresponding answers as entries matching the pairs (student and task).

The most important realisation I had was how much work it would take to conduct analysis on a data set such as this. Although I had worked with pedagogical data in the past, I only had experience managing complete or at least nearly complete data (i.e., data with very few missing data points). I realised that in many cases, when working with data from online learning environments, the data would and should resemble the data I had extracted. I had just not realised in advance that the data would be so sparse. In fact, I had long had the idea that a machine learning process or method could help me in this endeavour.

To establish any grounds to carry out future diagnostic processes, I had to be able to analyse these data. Together with Prof. Mørup, I discussed the possibilities of a clustering algorithm (unsupervised learning) for data analysis. Working with Prof. Mørup, I realised that we do, in fact, have a method for managing pedagogical data, such as data from online learning environments. Leveraging the knowledge of experienced data scientists, I realised that methods, such as the one utilised in Paper E (a co-clustering model called MIRM), can complete the data set by estimating the remaining 97.5% of the points and dividing the students and tasks into clusters. Prof. Mørup discussed with me the possibilities of finding clusters, and he anticipated that we would be able to find clusters with only 2.5% of the complete data set observed.

I want to briefly explain why co-clustering is an attractive choice in this context. Of course, I could have attempted to categorise the students based on regular statistics following ideas such as correctness and other prominent codes or created intervals together with measurement of how many tasks were answered. I am sure that some idea exists where students are only analysed or categorised at the classroom level based on the codes. However, for this project, I wanted to explore the possibilities for the whole data set.

In Paper E, we established and analysed these groups (clusters) of students and tasks to determine what interesting findings such co-clustering could present. It is important to note that when applying a model such as the co-clustering model, we do not have control over the clusters found. This makes the clusters very interesting to observe, since the model seeks to discover the possible signals from the data alone. In the case of this data set, the goal of the model was to generate co-clusters, which is a subset of rows that exhibit similar behaviours across a subset of columns, and vice versa.

I realised that this co-clustering model we trained on the data set could create the data points in the test set, which was left out of the training of each of the 25 models. Furthermore, this yielded better than baseline parameters, such as most common answer, most common student answer and most common answer for any given task. This provides researchers studying student answers, task design and feedback offered by online learning environments extended possibilities for extracting and collecting data. This co-clustering method could, based on merely 2.5% of a data set observed, create groups of students and tasks simultaneously, resulting in 7 groups of students and 395 groups of tasks (5 groups of students and 200 clusters of tasks using the subset). During this process of finding the best model, we learned to co-cluster these data so that the student groups were rather unpredictable, or what is referred to as a data set containing a lot of noise, and the model was therefore carried out in two settings—one where a student only had to have answered a single task to be included in the data set and another where a student had to have answered at least 100 tasks to be included. In the second setting, using the model on the subset of data, we found much more stable groups of students and effectively only 4 this time (the last group consisted of a single student) and only 200 groups of tasks.

4.2.3 Hypothesis 2.3

A diagnostic value can be interpreted by utilising the extensive amounts of data insights provided by applying unsupervised learning methods to data extracted from MatematikFessor.

The results of the co-clustering model presented in Paper E yielded several realisations regarding procedures for establishing grounds for carrying out diagnostics. I realised that students are rather unstable in their belonging to the groups we defined. Additionally the student groups initially only seemed to corresponding to some level of success in solving equations. However when analysing the construction of the student groups, in terms of prevalent error in each of them, we found that there were more to them than simply a level of success in solving equations. Importantly, I realised that the exploration of this hypothesis cannot be finished within the scope of this PhD project.

In total, we found 394 clusters of tasks. For a task to end up in a cluster with other tasks, they must share some characteristic in the form of answers provided. In this case, the goal is to provoke similar behaviours among the student population. Though many clusters of tasks were found, only 20 clusters included 7 or more tasks, and only 9 clusters had 10 or more tasks. Many of the clusters contained only a single task.

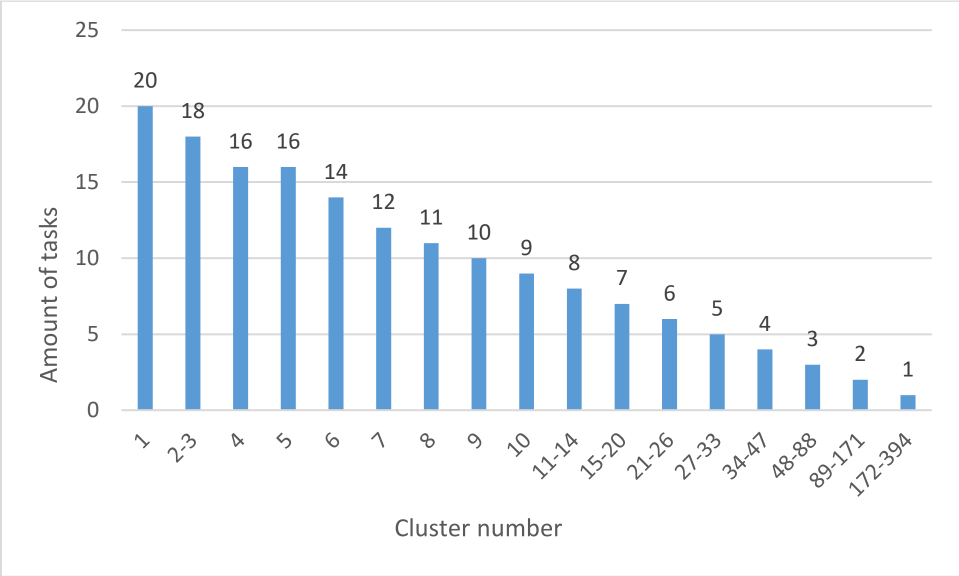


Figure 8: Bar plot of the distribution of tasks in clusters based on the original (full) data set.

When running the co-clustering with the data set excluding students who did not answer at least 100 tasks, we got a different result. In Paper E, we mostly concentrated the analysis on the full data set; however, for this *kappa*, I wanted to add some perspective and thoughts. After implementing this criterion, we were down to 2169 students, but every task of the 892 included in the design was still answered at least once.

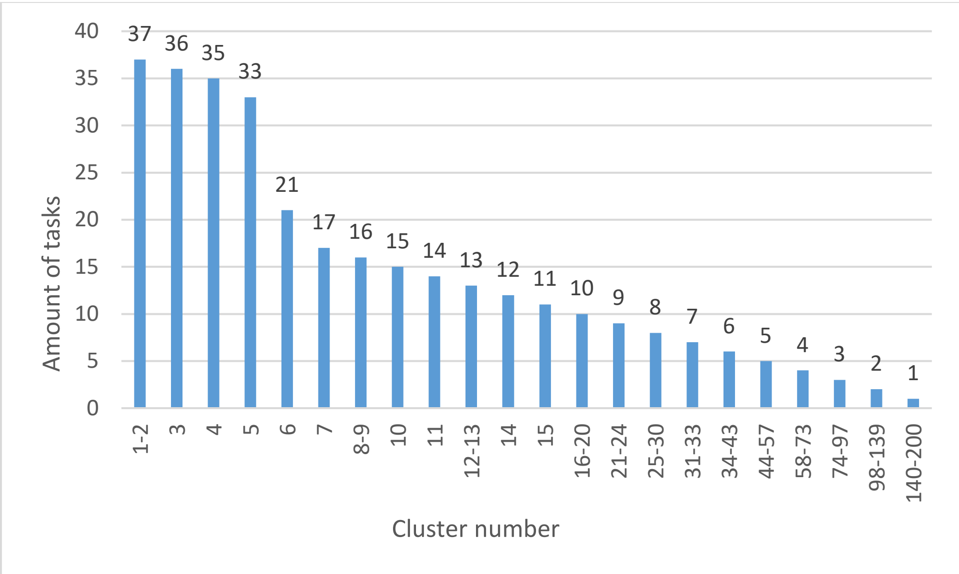


Figure 9: Bar plot of the distribution of tasks in clusters based on the subset where students had answered a minimum of 100 tasks.

One key difference in running the same method of co-clustering is that we received only half the number of clusters of tasks, which means that we had the opportunity to observe a setting where more tasks could be recognised as reflecting similar performances when observing the reason for an erroneous answer. This also means that we could observe a somehow stronger signal in these clusters when observing the differences within them. The potential for using these clusters of tasks as indicators for what tasks in particular work great for observing specific difficulties increases with a data set that has more observations.

However, I do not want to say that a partially observed data set is always worse than a more observed one. In this case, yes, we did find better clusters of tasks in terms of the grouping of difficulties, and we did also find groups of students who were less uncertain. With that said, there is also value to be found in efficiency or effectiveness in discovering tendencies and features in a very partially observed data set.

Table 16: Distribution of error types in % in the first four student clusters yielded by the model trained on the full data set, leaving out the last three clusters consisting only of 14 students

	c	u	n	o	r	rn	rm	av	mv	ra
C1	81.42	7.01	4.34	1.17	1.46	1.51	0.94	0.18	0.23	0.30
C2	92.49	2.60	2.08	0.48	0.48	0.6	0.43	0.06	0.01	0.12
C3	64.38	16.76	5.75	2.36	2.84	2.27	1.43	0.60	0.50	0.06
C4	24.74	55.82	4.34	3.63	2.21	2.21	0.92	1.08	0.42	1.13

	e	avrn	v	no	d	mvrn	mvr	zm	mvn	z	rnm
C1	0.26	0.18	0.14	0.08	0.02	0.19	0.00	0.03	0.06	0.07	0.04
C2	0.12	0.05	0.06	0.01	0.09	0.03	0.00	0.01	0.02	0.02	0.00
C3	0.55	0.36	0.24	0.18	0.28	0.22	0.00	0.03	0.08	0.05	0.08
C4	0.63	0.75	0.63	0.38	0.17	0.21	0.04	0.04	0.04	0.04	0.21

Table 17: Distribution of error types in % in the first four student clusters yielded by the model trained on the subset of the data, leaving out the last cluster only consisting of a single student

	c	u	n	o	r	rn	rm	av	mv	ra
C1	89.71	3.60	2.51	0.85	0.88	0.92	0.48	0.01	0.01	0.15
C2	74.08	10.65	4.83	2.30	2.60	1.83	0.81	0.56	0.04	0.39
C3	52.17	29.53	3.32	4.44	2.87	1.57	0.79	1.78	0.74	0.54
C4	18.03	68.09	2.18	5.76	0.14	0.60	0.54	1.84	0.42	0.21

	e	avrn	v	no	d	mvrn	mvr	zm	mvn	z	rnm
C1	0.14	0.07	0.08	0.04	0.01	0.05	0.00	0.01	0.02	0.02	0.00
C2	0.37	0.25	0.19	0.14	0.11	0.11	0.05	0.05	0.04	0.03	0.03
C3	0.56	0.39	0.23	0.03	0.08	0.12	0.18	0.08	0.04	0.02	0.02
C4	0.10	0.14	0.28	0.21	0.00	0.00	0.12	0.00	0.00	0.04	0.00

In Paper E, we also applied the MIRM to a subset of the data, where each student had provided an answer to at least 100 of the tasks. The model trained on a subset of the data validated the model trained on the large data set in such a way that we expect that with a tighter or more observed data set, we would be able to observe something more specific (or better) about the clusters. The model trained on the subset could co-cluster the students and tasks more precisely than the model trained on the large data set, since the data contained less noise. More precisely, the model presents a larger possibility of being correct in the placement of the items (students and tasks). The large data set contains more noise than the smaller one, in the sense that there are some students in the large (original) data set who answered very few tasks. It makes sense that these students are difficult to place, since we know very little about them. The good thing about the large model is that it can place a student in a cluster when that person has answered only a few tasks because we have such a large amount of data.

Regarding the composition the clusters, we would also expect that the model ran on the small data set is better able to place the students and the tasks. Something else to note is that on the student side of things, the model trained on the smaller subset did not form significantly different clusters. In comparing the distribution of the two models, we could not observe large differences in the distribution of correct percentages or difficulties. However, if we observe the tasks side of the matter, we can glean a more information from analysing these clusters. With the expected result of some tighter clusters, we can put ourselves in a position to better observe groups of tasks that together reveal a specific difficulty.

The diagnostic process would then be based on the idea that students in Denmark would be placed into a category according to one of the four clusters yielded by the model trained on the large data set, representing the variety of students as appropriately as possible. Within each of the clusters, it is possible to extract concrete tasks and tasks that serve as the best means for provoking a certain behaviour resulting in a specific error code. For example, we might look at students from Cluster 1. These students might on the surface be really good at solving equations and might be associated with results of having answered 80–90% of equations correctly. However, Cluster 1 also indicates that a student belonging to it should be struggling with equations that provoke the ‘negativity issue’ (answering with the additive inverse) resulting in the students answering the problem with the additive inverse. Analysing the clusters of tasks found in the subset, we see that by sorting the clusters first on a percent-wise distribution based on the error codes and then by the highest value of the ‘negativity issue’, we get an opportunity to locate tasks particularly good at provoking certain errors (see Table 18).

Table 18: Subsection of task clusters sorted to locate the clusters that are most likely to cause a negativity issue

	c in %	u in %	n in %	o in %	r in %	Total %
C 93	49	6	45	0	0	100
C 94	49	16	33	0	0	98
C 185	66	5	27	0	1	99
C 200	60	8	26	6	0	100
C 88	65	7	23	0	0	95
C 141	53	19	23	0	4	99
C 120	73	5	21	1	0	90

Noting the distribution of answers within the clusters allows us to observe what we can characterise as diagnostic value in this context. Looking at the tasks from Cluster 93 (the first row in Table 18), we see that 45% of the answers provided were incorrect due to the ‘negativity issue’. Equally important is that the remaining answers are distributed among very few other codes, in this case 49% ‘correct’ and 6% ‘unknown reason’. The particular tasks in Cluster 93 are as follows:

- $-3x = -15$ (type 4)
- $-4 - x = 9$ (type 3)
- $12x + 57 = 2x + 27$ (type ‘Algebraic’)

We did exemplify this in Paper E; however, for this *kappa*, I took the opportunity to present additional data and perspectives from the model trained on the subset that provides better options for choosing clusters relevant to the most prominent error codes. In terms of diagnostic value, this was another opportunity for it to be exemplified. When we can discover groups of tasks, such as those from Cluster 93, we can locate tasks and verify their diagnostic value, and this is true regardless of what type of equations they are. The diagnostic value of the tasks from Cluster 93 is therefore quite high based on the idea that the set of possible error codes is proportionally rather limited; at the same time, a large proportion of the answers are interpretable.

I realise that categorising students based on knowledge or, in this case, equation-solving ability or knowledge of how to solve equations could be considered an impossible or even unfair thing to do. Intuitively, I could have predicted that students would be very unstable or inconsistent in their answering of tasks. This is also what we observed. In the a posteriori analysis, I return to the discussion or the consequences of categorising students into groups based on performance solving linear equations.

4.2.4 Hypothesis 2.4

The task $8 + 4 = \_\_ + 5$ (Falkner et al., 1999) holds high diagnostic value and will serve as a good measure for comparing diagnostic value related to understanding the role of the equals sign.

This hypothesis is the product of the idea that tasks with diagnostic purposes related to the concept of linear equations could be represented differently than regular tasks, as they are from the online learning environment. When visiting Prof. Hodgen at UCL, I became familiar with the difficult process of designing tasks to evaluate students' knowledge or capabilities. During his work with the CSMS, Prof. Küchemann had formulated several tasks aimed at measuring students' difficulties conceptualising the rules and elements of algebra. Preparing to work with Prof. Hodgen and Prof. Küchemann, I reviewed the literature on students' difficulties working with equations and found that many of these studies also reported findings of utilising tasks.

As a part of my change in environment by visiting UCL, I worked under the supervision of Prof. Hodgen and Prof. Küchemann and designed a series of tasks supposed to function as diagnostic tasks. These tasks should be able to address a specific difficulty about a concept in relation to the concept of linear equations. These tasks should also be present in various representations, for instance, equality, when evaluating both students' understanding of that concept and their performance.

We ended up designing three 'ten-minute tests', as I named them. The idea was for the tasks to function through the online learning environment. I conducted a small pre-test featuring some of the tasks in two classrooms at two different schools using pen and paper before deciding on the 21 tasks going into each test. The tasks in the three tests were not unique to the tests. The tasks were mostly variations of each other, and some were identical. Included in these tests was the famous task by Falkner et al. (1999): $8 + 4 = \_\_ + 5$. I was fascinated with the idea that this task would only produce one of three different answers—7, 12 or 17—and these answers do all signify an interpretable and well-defined error. I kept imagining that this task is a very good diagnostic task. I decided to design variations of this task for comparison in the other ten-minute tests. This led to several ideas.

First, I wanted to use the original task in a replication study. Second, I wanted to prove that the original task is in fact a very good diagnostic task or rather a task holding high diagnostic value when relevant. Writing Paper D, we realised that the performance the authors found was rather extreme. In their reports, not a single 6^{th} grade student provided the correct answer. In our finings, 6^{th} grade students averaged a little over 60% correct answers and provided an answer sample space with the same values as the original study. I want to verify that my data were not somehow corrupted. Therefore, I conducted a small test in a single 6^{th} grade classroom and got the same results as I had gotten using MatematikFessor. The two studies, Falkner et at. (1999) and ours in Paper D, are 20 years apart, and there could be several factors explaining why this difference in the distribution of answers is so enormous.

Second, and perhaps most importantly, we investigated the diagnostic potential of the task. The results we got were that we could easily recreate the same 'feeling' in the task, in the sense that it was mainly the students' interpretation of the role of the equals sign that mattered in informing what answers they would give. However, we argue that it makes the most sense to formulate the equation with the empty space in the position right after the equals sign, since the variations we tested out in Paper D yielded relatively worse performance.

4.3 Hypotheses related to Research Question 3

I now present the posed hypotheses related to Research Question 3 and what came out of exploring them. In each section, I remind the reader of the related hypotheses, which have to do with possibilities regarding establishing interventions that can help address difficulties already detected or in working with students to address generic issues when solving equations.

4.3.1 Hypothesis 3.1

It is possible to set up an alternative (VR) environment in which teachers and students can explore equation solving while utilising metaphors to help guide students' equation-solving schemes.

During the exploration of this hypothesis, I went over a variety of options, since Lui Thomsen and I decided to build a virtual environment from scratch. The initial idea was that we would generate an escape room setting where the user (the student) would need to solve a puzzle involving solving or working with equations to exit the room. We wanted the room to enable exploration, requiring students to acquire the pieces needed to solve the puzzle. The main puzzle should consist of a balance model that was out of balance and advancing the clear purpose that balancing the model would solve the puzzle.

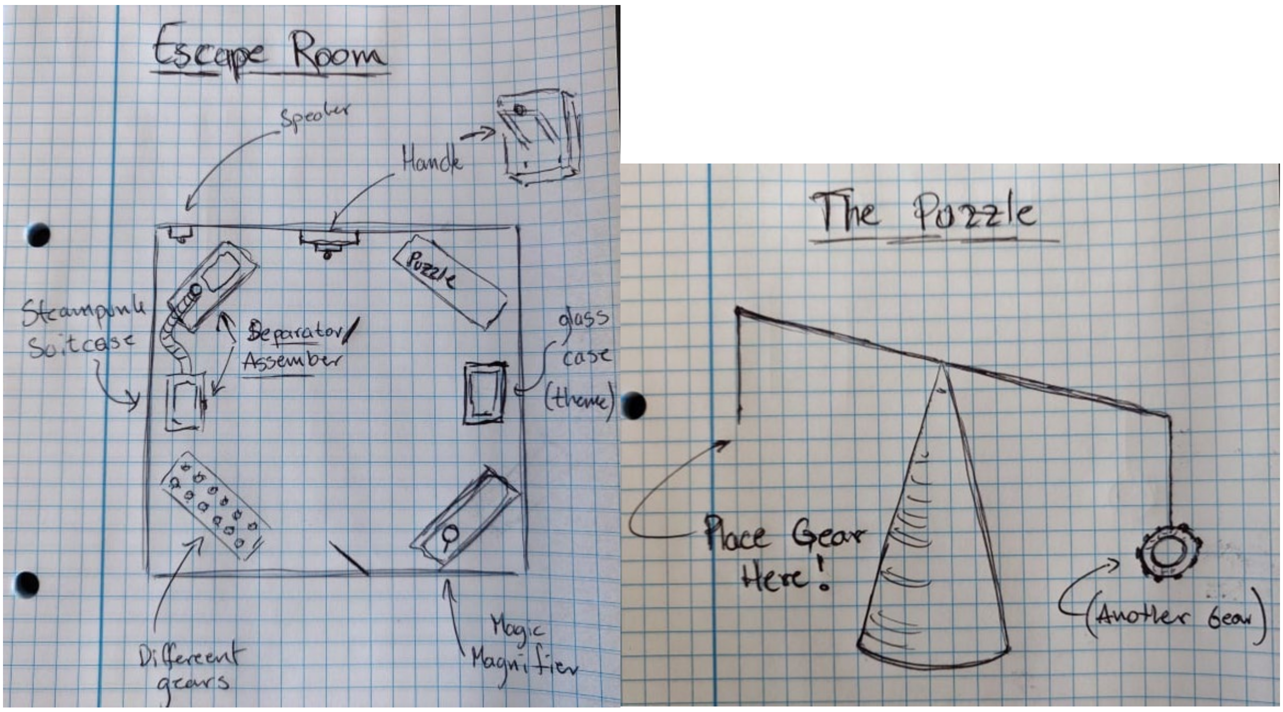


Figure 10: Initial escape room idea.

The idea was that the user would need to explore the room to find or construct the gear (the unknown) that would balance the model (see Figure 10). The user must use a magical magnifying glass to find the gear(s) needed to counterbalance the unknown. We also wanted the puzzles to get harder by requiring the user to construct a solution by using a combination of gears that make up a steampunk machine called the 'separator' or 'assembler' found in the opposite corner of the magnifying glass. This machine should let students transform the objects (the gears) by multiplying their weight or quantity to construct a solution to the puzzle. We worked with different objects that could represent numbers in the equations represented by the balance, and at some point, we switched to bottles instead of gears. This initial idea we actually designed and developed to the point that we did initial testing with a small group of students. This initial version incorporated a fully functioning virtual balance. The puzzles were to consist of a balance prepared with objects (bottles; see Figure 11) that do not balance the model. Scattered across the room, we had bottles for the user to find and use to solve the puzzle by achieving balance.

Figure 11: Balance setup in original puzzle for escape room.

As an additional exploration element, we implemented the feature of the magical magnifying glass, which gave the user the opportunity to see the hidden value of the bottles found inside each bottle cap (see Figure 12).

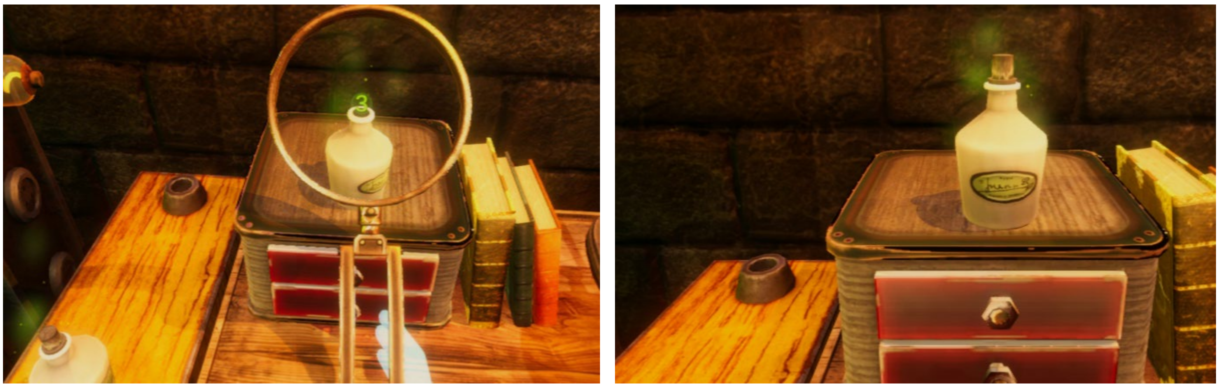


Figure 12: Magical magnifying glass.

While working with the mechanics and dynamics of this escape room, we realised that we were not quite clear on how students would actually learn about equations and equation solving. We thought the puzzle aspect was fun and entertaining; however, we found it difficult to come up with robust ideas that would simulate equation solving for lower secondary school students working with known sources of difficulties. We realised that we needed to be more in control of the elements, as the equations from the escape room were easily solved by the test students, and the rest consisted of looking for bottles that would fit this simple solution. We tried to think of additional objects that would provide the idea with more relevance for teaching linear equations. For example, we thought of manipulating the assembler and separator, as we wanted to keep the gamified feeling of exploring equations through this escape room.

However, we ended up returning to the idea that the VR game should incorporate exploration and an intervention against a known difficulty students experience with equation solving. Because we realised that the functionality of the virtual environment would let us invert the direction of the gravitational force, we thought we could provide new possibilities for working with the balance model in the classroom. We decided that negative numbers was the obvious choice, since this corresponded well with the idea of inverting gravity. We abandoned the more gamified idea of the escape room and decided that we were designing a teaching tool. Many of the realisations that we had after this decision to change directions regarding modifying the balance in a teaching experience we discuss in Paper F. We felt that we could incorporate the affordances (Otten et al., 2019) for teaching within the model and actually mitigate or avoid some of the constraints of the model that made it unfit for teaching equation solving (Pirie & Martin, 1997; Vlassis, 2002).

Working with the VR application, we had the opportunity to perform a study resembling a teaching experiment with 10 lower secondary school students. These students and their teachers had not been working with equations for long and thought it would be fun to participate in an experiment involving VR. This study is also presented in Paper F. We realised through the process of designing and developing the VR teaching tool we called ‘Equations Lab’ that it was difficult anticipating how the students would interact with the mechanics, dynamic and aesthetics (the MDA framework from traditional game design) in our environment. The underlying thoughts regarding the development process are not reflected in Paper F, since they were not fitting for a research paper format. However, it took us the good part of half a year to develop the application using a development platform called Unity.[5] All the 3D assets (objects) were modelled using free 3D modelling software called Blender.[6] I modelled and hand painted the assets in Blender in an attempt to achieve what we felt was the perfect setting for solving linear equations in VR. The process of developing the experience became a very personal project, and I believe that the outcome was better because we made everything from scratch.

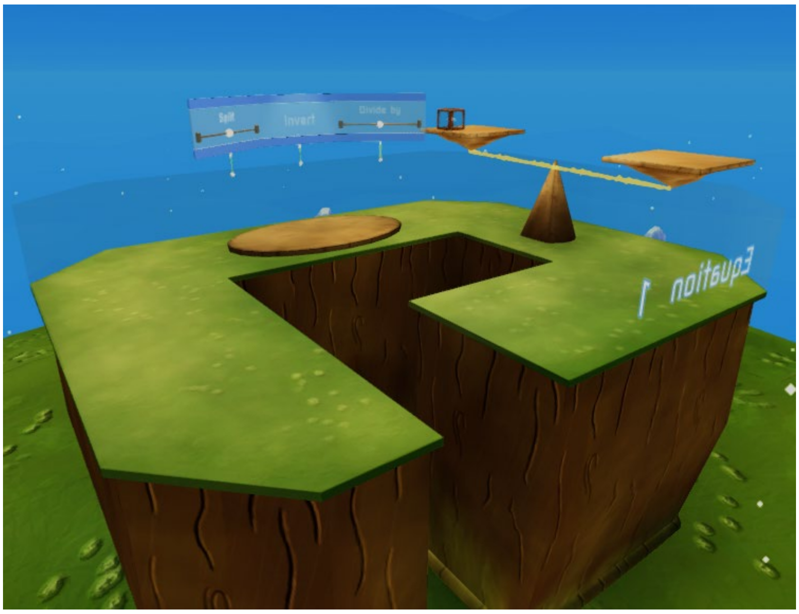


Figure 13: VR environment for teaching linear equation solving

Through Paper F, we managed to provide the students who participated in the study a unique experience, and we somehow had an epiphany working with the *invert* mechanic. The students had not seen such a term used before, and their schemes for working with equations in this experience seemed to assimilate towards the theorem-in-action that inverting a term and placing it on the opposite balance pan would maintain/restore

[5] Unity is a platform for developing digital games and experiences. Read more at https://unity.com/
[6] Blender is 3D modelling software for creating 3D assets, movies and more. Read more at https://www.blender.org/

balance. Their thoughts seemed to conform to the idea that instead of pressing down on one side, you can achieve the same effect of pulling up on the opposite side when working with the balance model. We did realise that connecting the scheme(s) for working with equations in VR and their already established schemes for solving (or working with) linear equations was not a straightforward process. I do not know whether we expected it to be straightforward, but I remember dealing with some level of disappointment from how successful the knowledge transfer to pen and paper was, since we were rather excited about our innovation.

4.3.2 Hypothesis 3.2

Utilising large amounts of data in addition to extensive coding and co-clustering of answers to tasks involving equation solving will pave the way for formative assessment opportunities, allowing teachers to address students' difficulties in solving linear equations using MatematikFessor.

Realising that I would not be able to present anything near a fully functioning commercial product that can provide better and more relevant feedback for teachers working with MatematikFessor, I instead made some important realisations and observations along the way. Hypothesis 3.2 was built on the idea of developing diagnostic insights via co-clustering and data analyses enabling the design of improved automated feedback for teachers using MatematikFessor. For that reason, this hypothesis builds on the previous hypotheses, 2.1 and 2.2.

When analysing the most relevant part of the distribution of codes, we can conclude that the formative tool could be established based on the classroom distribution of students matching the clusters of students found. As I explained in relation to Hypothesis 2.2 and Paper E, I would attempt to notify teachers of the distribution of students relative to the clusters found.

From Cluster 1, we can see that the most prominent errors were due to 'negativity errors'. Therefore, an automated feature offered by MatematikFessor could be to suggest tasks from the literature related to this problem and tasks found in this data set by the co-clustering algorithm that are particularly well suited for provoking this error. Additionally, the automated system could provide teachers with the means to teach the important features and common misunderstandings when interpreting or solving such equations.

As an example, we could look at a task from my collection of equations that seems to be well suited for provoking the negativity issue, specifically $-3x = -15$ (type 4 (2784)). When working with this equation, students should be made aware and learn about the different features and interpretations of the minus sign (Gallardo, 2002). Additionally, we are aware of equations similar to this one from the research literature, such as the famous task $-6x = 24$ (Vlassis, 2002). Thus, we are familiar with equations that are detached from a model and cannot meaningfully be represented on, for instance, a balance model.

As another example, students from Clusters 2 and 3 might occasionally make errors such as 'rearranging for sense making'. The tasks that are particularly good at provoking this error are $4 - x = 5$ (type 3) and $3 - x = 6$ (type 3). Here, we see two similar tasks in the form $a - x = b$, where b is larger than a, and thus, x is a negative number. For these equations, the most common answers were as follows:

$\mathbf{4 - x = 5}$		
Answer	#Answers	Code
-1	7066	c
9	4212	r
1	1296	n
-9	391	rn
6	126	u

$\mathbf{3 - x = 6}$		
Answer	#Answers	Code
9	3804	r
-3	2729	c
3	1692	n
6	189	u
-9	187	rn

With both these cases, the code 'r' covers a large number of the answers to these two tasks. In fact, in the latter case, we can see an extremely rare case of the correct answers not being the most popular choice.

I believe that the data sets derived from online learning environments may initially look very different from data sets gathered by going into a classroom with a set of tasks suitable for a single intervention. What we learned from extracting data from MatematikFessor is that a large amount and variety of students access the tasks within it.

I realise that using the clustering method to create the exact means for improved formative assessments might be clumsy or too unspecific. However, I believe that such studies can enhance knowledge about how students behave when solving linear equations in online learning environments and how we can improve assessment and task design principles. I believe that applying a coding system such as the one I have presented in this *kappa* can help establish that online learning environments can, on individual student basis, create track records of the codes students collect while solving tasks. These codes can then be reported to teachers so that they know what types of equations students struggle with and why. This would serve as a more individualised and perhaps precise and fair form of feedback for teachers.

I do believe that the coding system could enable a better outset for statistical overviews provided by online learning environments, such as MatematikFessor. By combining the statistical tools already present and the coding from Paper E, these tools will gain new powerful abilities in providing feedback for teachers and students. However, I realised through working with the coding system that we might need to revise it, since the distribution showed that some error types were only minimally relevant. Nevertheless, through the co-clustering study with the MIRM model, we can locate groups of tasks that perform as we desire even across types of equations.

Chapter 5 A posteriori analysis and validation

In this fifth chapter of the *kappa*, I intend to revisit the methodological structure, engaging in a posteriori analysis and validation part of the DE structure, which should be organised in terms of contrast with the a priori analysis. During this phase, the data collected, realisations and observations made are analysed to identify convergences and divergences in relation to the preliminary analyses. The hypotheses made during the conception phase are put to the test during the validation and typically involves multiple data sources. Artigue (2015) argued that the validation process should not impose a perfect match between the two analyses. She further mentioned that the methods and tools for comparing the preliminary analyses and the a posteriori analyses are constantly evolving.

5.1 Discussion of findings in terms of hypotheses set during the a priori analysis

In this section, I return to what was established in the preliminary analyses and discuss significant convergences and divergences and how to interpret these. For this exercise, I also want to include the research questions as a point of reference and an important factor in the discussion.

5.1.1 Hypotheses related to Research Question 1

In the a priori analysis, I established four hypotheses related to Research Question 1:

- 1.1. In MatematikFessor's database, there are already thousands of tasks about linear equations relevant for lower secondary school students. These tasks have enough diagnostic potential in terms of design to detect students experiencing difficulties.
- 1.2. Designing a collection of tasks with the aim of uncovering the difficulties found in the literature can strengthen the possibility of achieving a set of fundamental tasks with significant diagnostic potential in detecting students experiencing difficulties.
- 1.3. Choosing input fields instead of multiple-choice options when selecting/designing task involving solving linear equations will enable opportunities for discussing diagnostic potential, future design principles and knowledge on proper distractors.
- 1.4. The four aspects that define the scheme as a part of the TCF can serve as a comprehensive means for designing tasks to ensure diagnostic potential in an assessment situation incorporating an entire action.

I established in Paper A as part of the preliminary analyses and as a prerequisite for the design that the tasks already implemented in MatematikFessor were ill suited for establishing solid grounds for appropriate and valid diagnostics. In other words, to the best of my knowledge, I can say that Hypothesis 1.1 should be rejected. On the contrary, having learned from the research literature both about learners' difficulties solving linear equations and task design principles, Hypothesis 1.2 should be accepted. I mainly base this acceptance on the analysis conducted when exploring the later hypotheses; however, since I formulated that the tasks should have diagnostic potential, I must accept Hypothesis 1.2 based on the knowledge acquired through the literature review conducted as part of Paper B.

Choosing input fields instead of resorting to multiple choice was, in my opinion, the right choice. Even though the data set became noisier due to the many uninterpretable answers students gave, the trade-off would, in my opinion, be a lack of knowledge about the students and especially the tasks. I discovered something about the tasks because I designed them using an input field rather than multiple choice. The 'one more/one less' error type is a significant finding and would not have been discovered had I used multiple choice. I accept

Hypothesis 1.3, since the discovery of the ‘one more/one less’ error type verifies that I could not have created meaningfully distractors based on the findings from Paper B or other expert knowledge. There is more to it than just accepting Hypothesis 1.3 based on this finding alone. However, I firmly believe that exploring the potential gains in using input field instead of multiple-choice have extended my knowledge on the answers that would be provided in MatematikFessor and thereby a reason for establishing new and more refined design principles for tasks about equations for the environment.

Accepting or rejecting Hypothesis 1.4 is a matter of accepting an idea or a design philosophy. I think the idea presented in Paper D is valid, and I think that creating or designing learning trajectories or sequences with the aspects of the action (the scheme) is an idea that creates awareness about how difficulties can arise when students solve mathematical tasks and engage in mathematical situations.

5.1.2 Hypotheses related to Research Question 2

In the a priori analysis, I established four hypotheses related to Research Question 2:

- 2.1: Extensive didactic coding of answers to tasks can help create a means for exploring the extent of different reasons for students’ erroneous answers based on answers to linear equations from MatematikFessor.
- 2.2: Unsupervised learning methods (specifically MIRM) together with additional didactic coding of answers to tasks can help create a means for categorising students based on answers to linear equations from MatematikFessor.
- 2.3: A diagnostic value can be interpreted by utilising the extensive amounts of data insights provided by applying unsupervised learning methods to data extracted from MatematikFessor.
- 2.4: The task $8 + 4 = \_\_ + 5$ (Falkner et al., 1999) holds high diagnostic value and will serve as a good measure for comparing diagnostic value related to understanding the role of the equals sign.

The choice to apply the codes is discussed in Paper E, and the idea is mostly presented as going beyond correctness. The findings presented in this *kappa* and in Paper E gave me an extended perspective on the possible answers students provide and that a coding system such as the one applied in this context can collect different erroneous answers into the same category of reasons for that particular error. To explore Hypothesis 2.2, I applied co-clustering in Paper E to attempt to categorise students as equation solvers. From my perspective, the groups the students were placed in did provide some information. The groups were very unstable in the sense that the placement of the students in the cluster was quite uncertain. Second, the differences in behaviours that led to the groupings were not solely based on interpretable errors; rather, they were based on uninterpretable errors and the proportion of correct answers. However, in Hypothesis 2.3, I established that the information we can extract from the co-clustering of the tasks is quite applicable. The idea that we can extract precise behaviours from different types of tasks and their probability of collecting only certain types of errors is applicable to task design and assessments for feedback in MatematikFessor.

The last hypothesis (Hypothesis 2.4) served as an opportunity to exemplify diagnostic value and to test what is considered a good task in a new environment to observe the outcome. The results showed that the task was still equally effective in collecting only a certain set of answers, and we showed with the variations that the placement of the blank space is important and is a testimony to how good the task is. Additionally, we found through the replication study (Paper D) that the distribution of answers revealed results that were far from the truth when applied to Danish students in MatematikFessor 20 years after the original study.

5.1.3 Hypotheses related to Research Question 3

In the a priori analysis, I established two hypotheses related to Research Question 3:

- 3.1. It is possible to set up an alternative (VR) environment in which teachers and students can explore equation solving while utilising metaphors to help guide students' equation-solving schemes.
- 3.2. Utilising large amounts of data in addition to extensive coding and co-clustering of answers to tasks involving equation solving will pave the way for formative assessment opportunities, allowing teachers to address students' difficulties in solving linear equations using MatematikFessor.

In Paper F, we presented our attempt at designing, developing and testing a novel VR application for teaching equation solving with a focus on exploring the difficulties related to solving equations made abstract by negativity and negative numbers. We found that students could pick up on the idea that negative values affected the balance in a different way than positive values. However, our idea did not incorporate any means for teaching the idea of the missing operator on the balance. The behaviour we saw from the students allowed us to infer the idea that the operations in standard algebraic equations caused them difficulties, and the VR balance teaching technique did not account for these difficulties.

Hypothesis 3.2 has to do with the idea of establishing an improved, automated way to assess for students experiencing difficulties solving equations in MatematikFessor. Based on the knowledge gathered through exploring the hypotheses in this dissertation, I believe that automated assessment in online learning environments can be improved. Applying the codes to a select set of equations based on the clusters and the diagnostic value of the tasks could be a first step. Statistical overviews and reports that contain classifications of students based on their interactions with these select equations should be the next novel attempts at improving automated assessment in environments such as MatematikFessor.

5.2 Discussion of findings in terms of contrast to the preliminary analyses

For clarification, this section might not be prescribed by the DE framework or structure. However, as this process was applied as part of the structure of the *kappa*, and the fact that it was applied in retrospect, I find it appropriate to assess for convergences and divergences in these analyses as well.

5.2.1 Convergences and divergences from the didactical analysis

An important part of the didactical analysis was the establishment of students' difficulties in learning about or working with linear equations and linear equation solving. Much of this was learned through the literature review presented in Paper B. In the didactical analysis, I attempted to establish what prior research could reveal in terms of the teaching and learning of the concept of linear equations and linear equation solving. Within the didactical analysis, I also mentioned that I, to some extent, wanted to stick with the concept of didactical situations (Brousseau, 1997) and the emergence of such in online learning environments. Relying on the difficulties found in the literature and the idea of a terminology defining the concept of what I mean by difficulties, I wish to highlight and discuss the convergences and divergences in my realisations compared to the literature. In line with learning about didactical situations related to teaching and learning the concept of linear equations, I want to draw on several realisations.

First, I want to discuss what I realised when assessing the students' errors based on the data set I explored using the co-clustering model. We noticed several things that I did not expect based on the literature review. The most notable was that students in MatematikFessor ran into the obstacle of providing the wrong answer by presumably pressing the wrong button on the keyboard. I am referring to the error code 'o' (one more/one less). This became important in the further analysis of didactical situations for analysing students' difficulties in terms of gathering information based on which teachers and researchers can hypothesise about how to best move students forward. This error of pressing the wrong button when providing an answer in an environment such as MatematikFessor has to do with something very different from mathematical knowledge or schemes for solving linear equations. This error is connected to some affective value, such as a lack of concentration or focus due to factors that might prove very difficult to determine via MatematikFessor. On the other hand, variables, such as what time of day the student provided the answer or how fast the student provided the answer, might help determine the cause of the error. In my opinion, this issue is obviously not caused by difficulties of a mathematical nature, which is very important to note when analysing data from online learning environments.

The equations designed using the findings from the literature review are based on variations in their types. The types were chosen in an attempt to cover what might be considered reasonable for lower secondary school students. To that end, I believe that the design was rather successful. None of the types of equations presented in the design seemed to be too difficult, based on the findings from the data analysis that no type suffered more than 33% wrong answers.

Second, I wish to address the convergences revealed by the analysis of the large data set. As anticipated, the negative numbers or the minus operation proved to cause the students difficulties. In all the groups of students we identified in the co-clustering, we observed the negativity issue as most present only second to the unidentifiable or unknown issue.

Rhine et al. (2018) argued that issues such as 'acceptance of lack of closure' (referring to e.g., Collis, 1978) or the process-product dilemma (referring to the work of Davis, 1975; Sfard, 1991) are apparent in students' work with algebraic expressions and equations. Rhine et al. (2018) also connected this issue to the famous task $8 + 4 = \_\_ + 5$ (Falkner et al., 1999), and the results showed that no 6th grade student out of a sample of 145 was able to provide the correct answer of 7. However, in our replication study, a little over 60% of 6th grade students were able to correctly answer 7, and what is extremely important in this context is that I got the same answers as the original study. Perhaps concepts such as 'acceptance of lack of closure' and the 'process-product' dilemma cannot be connected to tasks such as $8 + 4 = \_\_ + 5$ in the way we thought. Perhaps 20 years of progress in teaching has taught teachers, students and other stakeholders to be aware of issues when interpreting the equals sign to an extent that we now see a different distribution in the answers. This would indicate a positive change in the distribution. However, I found that equations that were 'backwards' in formulation did attract a fair number of errors. Here I mean errors connected to the interpretation of the equals sign. What I saw in the analysis of the large data set was that the largest clusters of tasks, meaning the largest groups of tasks that were grouped due to similar performance, were tasks of the same type from the structure group grouped together or were the majority of the tasks considered 'backwards' in formulation. I expected tasks of the same type to be grouped together to a certain extent, especially if they called for the same types of errors and thereby the same error codes. What is more interesting is that these subtypes I refer to as backward variations of both Types 3 and 7 were grouped together more than any other identifiable task type.

5.2.2 Convergences and divergences from the epistemological analysis

In my exploration of the concept of linear equations, I focused on the idea of detecting students experiencing difficulties, diagnosing the origin of these difficulties and setting up the means for intervening against them. In the epistemological aspect of what linear equations are and have been, I attempted to explore different aspects of the concept of equality of expressions.

5.2.2.1 What is an equation and what is a solution to an equation?

Having worked mostly with an online learning environment, I have somewhat limited experience with working directly with the concept of equality. However, in Papers C, D and F, I explored equality and the equals sign from different angles related to what equations are and what we mean by solutions to them.

In most cases, we refer to equations rather than expressions when we want to identify something that has a possible solution. Sometimes, the distinction between expressions, equations and functional formulas can seem tricky and become a barrier or an obstacle when we want to explain what we mean or what is important in the current situation or setting. Prediger (2010) established that we apply a specific meaning to the equals sign when talking about equations, and asking for a condition under equality can be upheld by identifying the value of one or more unknowns. Similarly, Prof. Niss (personal conversation) identified that we can be talking about a certain set of situations where the relational aspect of equality acts like a proposition asking for a solution.

During the experiment with the VR application, I found that the balance model and the setting in which we used and modified it provided the anticipated idea of equality. The students were intuitively aware that some unknown ensured that equality between the two pans was in fact established and that a 'truth' or a solution could be determined (see Otten et al., 2019). This idea of the equals sign proposes that the possibility of upholding it is different when the equations are represented on the balance model. Here, we did not represent equations without solutions, even though these equations might not be present in the typical classroom either. My point is that in Paper F, students got to experience the properties of the equals sign in its relational sense where the system or the situation kept other epistemological obstacles in check, since the system or the application would not allow for misinterpretations of the role occupied by equals sign or letters of the same appearance having different numerical value, etc. However, this does not change the perception of the equals sign; rather, it requires us as mathematicians or students to be rather specific in interpreting or teaching the idea or concept of an equation when attempting to work out a solution. When working with equations, the perception or the meaning of the equals sign is added a prompt. Prof. Niss (personal conversation) proposed the idea that when attempting to solve equations, the unknown resembles the job applicant we as interviewers are recruiting to fill a needed position.

However, the representation on the balance to the symbolic representation did not achieve knowledge transfer as easily as we had hoped. With the balance in VR, we did not let students' misinterpretations of how mathematics works influence the choice of solution. Students might not know how and why they reached a solution using the tools and balance in VR. However, they could not reach an incorrect solution. I argue that this creates a unique opportunity to learn about both the concept of equations and their solutions in controlled and explorative situations.

In the task design using Vergnaud's aspects of the scheme (Paper C) featuring the dual scheme idea focusing on the concept of substituting equal terms, I learned first that designers should be careful when designing tasks. For teachers to hypothesise about their students' errors, we need very specific tasks or possibly sequences of tasks. Falkner et al.'s (1999) $8 + 4 = \_\_ + 7$ task is certainly a very good task for determining whether students are capable of reading expressions including an equals sign without interpreting the equals sign as a 'do something signal'. However, much more can be learned about students' interpretations of expressions where the equals sign plays a major role. In Paper C, I hypothesised about students' ability to interpret the equals sign or the structure of two expressions. The aim of writing this paper was to hypothesise about creating diagnostic potential or diagnostic situations detached from the necessity of having a conversation about the students' difficulties as part of a diagnostic process.

5.2.2.2 Epistemological obstacles

In this section, I go into more detail on the concrete errors students make while solving linear equations based on the explorations I conducted. The findings rely mostly on Papers B and E. In Paper B, I designed a set of linear equations that is supposedly a reasonably comprehensive set suited for lower secondary school. The linear equations were all based on the difficulties and obstacles identified in the literature review in the same paper. In Paper E, I summarised the four categories of difficulties in relation to the following:

- *The concept of numbers*—This involves negative numbers (specifically the role of the minus sign in expressions) and numbers in expressions that belong to sets in general that stretch beyond the natural numbers (i.e., rational numbers and zero) (e.g., Gallardo, 2002; Vlassis, 2002).
- *The equals sign and its role in expressions*—This issue is mainly concerned with how the equals sign is interpreted in concrete equations and how the structure of the expressions is thereby made sense of. In some cases, terms might be disregarded or misread (e.g., Kieran, 1981; Matthews et al., 2012; Prediger, 2010).
- *Strategies and transformations*—Depending on the complexity of the equations, different strategies might come into play, including different sorts of transformations and procedures. This includes conventional concepts, such as the procedures and roles of operators when rearranging or transforming expressions to reach a solution (e.g., Jankvist & Niss, 2015; Kieran, 1985; Linsell, 2009).
- *Letters in expressions*—These issues could be related to the role and handling of coefficients in terms of coefficients being added or multiplied onto the unknown (e.g., Küchemann, 1981).

None of these four categories are easily summarised or condensed into concrete erroneous actions students make while solving linear equations. The four categories rather referred to areas of attention. To discuss the findings from analysing students' answers to the equations designed based on the above ideas, I wish to compare my findings to a very specific primary source on students' errors: the study conducted by Kieran (1985). Based on this study's findings, I made the following tables (Table 19-Table 21). I find this an interesting comparison, since both the present study and Kieran (1985)'s study worked with approximately the same types of linear equations.

Table 19: Errors made by both novice and intermediate equation solvers (adapted from Kieran, 1985, p. 143):

Inversing a subtraction with a subtraction or failure to do so when necessary, e.g., solving $16x - 215 = 265$ by subtracting 215 from 265 or solving $37 - b = 18$ by adding 37 and 18.
Giving up when attempting to solve using the substitution procedure.
Inversing an addition with an addition, e.g., solving $30 = x + 7$ by adding 7 to 30.
Computing a coefficient with a non- coefficient, e.g., solving $2 \cdot c + 5 = l \cdot c + 8$ by adding 2 with 5 on the left side.
Forgetting that concatenation means multiplication, e.g., considering $6b = 24$ as $6 + b = 24$.

The first thing I want to highlight is the two issues common to both groups: inversing subtraction with subtraction and inversing addition with addition. In my coding, I had the codes 'rn' and 'ra'; these two codes referred to rearranging with a negativity issue and rearranging with an addition issue. In addition, I used the code 'r', which I call rearranging for sense making. With this code, I refer to instances where it might not be totally obvious how the student came to the solution or the conclusion that led to the solution. Furthermore, I discovered that I could find instances of students making similar rearrangements with multiplication and division. It is not only issues related to addition and subtraction that cause students to apply the inverse mechanism when attempting to solve linear equations by working backwards or applying transformations. Confusing multiplication and division was also very much present in the data related to the equations that enable it. What I found is that depending on the actual formulation or structure of the equations, I could find these rearrangements or transformations that did not follow the rules upholding equality. Additionally, I found that students answering with the additive inverse to the actual solution was an issue that persisted across all the equation types in my design. However, as I mentioned in the analysis and realisations of applying the error codes, this issue was probably a consequence of my design choices.

In this format or with the analytical tools I chose, I was not able to justify or detect whether a student gave up while solving an equation; hence, I was not able to detect whether a student gave up while using a substitution procedure. Neither was I able to detect whether a student did compute a coefficient with a non-coefficient, as in the example from Kieran (1985). I was able to detect that students under the right circumstances were willing to forget that concatenation meant multiplication. However, in instances where students were to solve equations such as $2x - 7 = 13$, this error of forgetting the meaning of concatenation was not present compared to other errors. We can learn from this that errors involving negative numbers and handling negativity in general are much more of a problem, at least among Danish students. Next, I move on to the errors made only by intermediate equation solvers (Kieran, 1985).

Table 20: Errors made only by intermediate equation solvers (adapted from Kieran, 1985, p. 144)

Leaving the unknown with a negative sign in front of it, e.g., $-x = -17$.
Changing an addition to a subtraction when transposing, but then commuting the subtraction, e.g., $30 = x + 7$ -> $7 - 30 = x$.
Transposing only the literal part of the term and leaving the coefficient behind, e.g., solving $7 \cdot c = c + 8$ by writing $7 - 8 = c : c$.
Dividing larger by smaller rather than respecting the order for inversing, e.g., $11x = 9 \rightarrow x = 11/9$.
Computational error involving positive and negative numbers.
Inversing a one-operation addition equation twice by inversing the addition and then dividing the unknown by the result of the subtraction, e.g., solving $n + 6 = 18$ by subtracting 6 from 18 and then attempting to divide *n* by 12.

Kieran (1985) found that negative numbers and the minus sign in general caused a lot of computational issues for students. In the analysis of my data, I was able to detect different issues of the improper handling of negative numbers and negativity. The second, third and fourth elements in the above table reflect something that is very similar to what I refer to as ‘rearranging for sense-making’ (my type ‘r’ error). The presence of this error in my data makes me believe that this error is in fact, alongside the errors ‘n’ and ‘o’, the most important. By important, I mean that such errors can be rather tricky to notice when simply observing the answers students give without observing their processes used to reach a solution.

Table 21: Errors made only by the novice equation solvers (adapted from Kieran, 1985, p. 144)

Not using the order of operations convention.
Not knowing how to start solving a given equation-type.
Inversing a multi-operation equation before collecting together the multiplicative terms.
Not using the convention that two occurrences of the same unknown are the same number.
Giving precedence to an addition when it is preceded by a subtraction.
Inversing a two-operation equation only once and then using the result of that operation as the solution.

The errors mentioned in Kieran (1985), which are made only by novices, are complicated for me to compare my results against. Working with many students through online learning environments offers the opportunity to gather many answers from many students to many tasks. However, what I am not able to observe is also important to notice. Linsell (2008, 2009a, 2009b) also worked with linear equations with reference to Kieran but with a focus on strategies for finding solutions. Some of what Kieran (1985) observed I would refer to as strategic or process structuring errors that are perhaps not observable or deducible from the solutions reached.

In many ways, I could verify that these errors Kieran found almost 40 years ago are still present today. Among all the errors that Kieran (1985) identified, I was not able to identify the issue of when students would give up after being unsuccessful using a substitution technique, nor did I aim to observe the students' choice in strategy or rule of action in the data from MatematikFessor. However, I was able to verify the issue where students would forget that concatenation means multiplication.

A difference and perhaps a strength of my study is that I was able to work with 892 different equations and gather data in the form of answers at the rate I did. Working with a reasonably comprehensive set of equations made it possible for me to distinguish the different kinds of errors and in some cases pair them with a certain type of equation(s). I imagine that this study can help other task designers working with linear equations to not only create distractors for working with the multiple-choice option but also in establishing diagnostic potential for their types of equations.

Building on the knowledge from the literature review and Kieran (1985), I can confirm that errors were not present for every type of equation. My study confirms that the type of equation must (and intuitively only should) afford certain types of erroneous behaviour. This finding underlines the idea that "you find what you are searching for". To prevent this idea from corrupting the emergence of new findings or discoveries, such as the error type 'one more/one less', I believe that it is necessary to attempt to make a reasonable comprehensive set or structure of tasks and let the data analysis reveal what erroneous behaviour can be learned.

5.2.3 Convergences and divergences from the institutional analysis

Coming to the last section of the a posteriori analysis, I review the findings against what I learned about online learning environments and their earlier use. I wish to compare my findings exploring the detection, diagnosis and intervention framework from the Maths Counsellor Programme in relation to linear equations against feedback systems as they are currently employed in online learning environments.

What I found in the small survey and review of teachers' use of feedback tools in MatematikFessor is that this feedback is based purely on statistics measuring right or wrong answers. In Utterberg Modén (2021), teachers would experience difficulties interpreting the feedback options from statistical dashboards representing students' progress.

Another thought that keeps coming back to me is the idea of letting students' teaching needs be deduced from their performance. By performance, I mean a measurement controlled by the proportion of correct answers relative to the number of tasks engaged with. I think we cheat ourselves out of learning from our students about our own teaching or about their previous knowledge or schemes used when engaging with (new) situations. Statements such as 'Students X had only a 64% proportion of correct answers when working with mixed linear equations, so I must teach Student X the idea of solving equations again' I believe are unhealthy and implement the 'big idea' of formative assessment mentioned in the introduction (Wiliam & Thompson, 2007). Furthermore, the idea that formative assessment can be built on this basis compares interestingly to the issues mentioned in Bennett (2011). As discussed in Paper C, Bennett (2011) raised several issues regarding the implementation or execution of formative assessments by teachers. With the coding system, I believe that we can actually move away from what Bennett refers to as his *definitional issue*: that when treated as an instrument, formative assessment is a test that produces a score from which teachers can deduct a 'diagnostic value'. Even though the codes can be viewed as just a score, they can offer so much more information than what the automatic scoring can offer in MatematikFessor at the moment. Feedback for teachers from an online learning environment is not to be confused with a test in that data are accumulated over time and hypotheses are continuously changing with student learning.

Acknowledging that formative assessment or assessment for learning is an idea that teachers are implementing in their teaching for themselves, we also need to be aware of what parameters make for fruitful formative assessment. I keep coming back to the question of how, if I begin with the second key strategy from Wiliam and Thompson (2007), to justify building a concept of diagnostics around performance based solely on a proportion of correct answers (possibly limited to a topic or a concept in mathematics). If so, how would I then justify following this interpretation after observing the error type 'one more/one less' and its propagation? Bennett (2011) claimed that formative assessment requires teachers to possess substantial knowledge and experience that enables them to make productive 'formative hypotheses' to act on. Building these hypotheses on a proportion of correct answers only enables teachers to hypothesise about what tasks or topics within mathematics students can or cannot do.

My recommendation for developers working with online learning environments and digital resources capable of automated scoring of answers to tasks is, based on this study, to redefine the statistical overviews and their basis for construction. I am sure that statistical overviews can serve a beneficial purpose in how teachers gather information about their students' mathematical capabilities. However, I am also sure that if the statistical overviews are based solely on correctness, these overviews will possibly drive teachers to think that the idea (in environments with automated scoring) is to reach a certain proportion of correct answers. I would like to challenge the idea that the understanding of a mathematical concept has a one-to-one correspondence with a certain threshold of a correct answer percentage.

As mentioned under the institutional analysis, a teacher made the comment stating that "I tell children and parents that the target is 90% correct when working on MatematikFessor". What I found during the analysis and especially the coding of the data is what I consider proof that we must not base statistical overviews in environments with automated scoring on a proportion of correct answers. Type 'o' errors (one more/one less) and the proportion of errors that were connected to them are to me proof that teachers building their ideas of students' understanding of equation solving on a proportion of correct answers can be misled.

Chapter 6 Concluding remarks

Working with the hypotheses and the preliminary analyses, in this final chapter, I wish to return to the research questions I set out to explore before discussing my reflections on the entire project. Afterwards, I will revisit the DE structure and the working questions I posed. The research questions were inspired by the framework established by Jankvist and Niss (2015), specifically the Maths Counsellor Programme's approach of *detecting* students experiencing difficulties, *diagnosing* the origin of these difficulties and creating *interventions* to hopefully move past such difficulties. This framework was applied to students working with linear equations in MatematikFessor. For each research question, I attempted to create a setting that allowed me to explore the question in both a broad and narrow sense.

6.1 Research Question 1

The first research question leaned into the idea of detecting students experiencing difficulties solving equations. The concept of task design was the focus when attempting to detect students experiencing difficulties in online learning environments. The main interaction with students was established by collecting data consisting of students' answers to tasks about linear equations. Research Question 1 was as follows:

What design principles are appropriate when structuring and designing tasks capable of detecting students facing mathematics-specific difficulties related to learning the concept of linear equations and equation solving?

a. *What principles are appropriate when exploring one specific difficulty (the equals sign)?*
b. *What principles are appropriate when exploring difficulties related to the more general topic of equation solving found in the literature?*

Working with the first research question in the broad sense (b), I conducted a literature review that would enable me to establish a set of design principles to create a reasonably exhaustive set of equations suitable for implementation in MatematikFessor. I established four overall categories of difficulties, a system of linear equations (starting with a linear equation in its most general form) and a set of design principles (mainly variation theory). The set of equations should, to a reasonable extent, be considered a set of fundamental situations (Brousseau, 1997). The four categories of difficulties were originally established in Paper B and later adapted in Paper E:

- *The concept of numbers*—This involves negative numbers (specifically the role of the minus sign in expressions) and numbers in expressions that belong to sets in general that stretch beyond the natural numbers (i.e., rational numbers and zero) (e.g., Gallardo, 2002; Vlassis, 2002).
- *The equals sign and its role in expressions*—This issue is mainly concerned with how the equals sign is interpreted in concrete equations and how the structure of the expressions is thereby made sense of. In some cases, terms might be disregarded or misread (e.g., Kieran, 1981; Matthews et al., 2012; Prediger, 2010).
- *Strategies and transformations*—Depending on the complexity of the equations, different strategies might come into play, including different sorts of transformations and procedures. This includes conventional concepts, such as the procedures and roles of operators when rearranging or transforming expressions to reach a solution (e.g., Jankvist & Niss, 2015; Kieran, 1985; Linsell, 2009).
- *Letters in expressions*—These issues could be related to the role and handling of coefficients in terms of coefficients being added or multiplied onto the unknown (e.g., Küchemann, 1981).

The design principles resulted in 892 linear equations distributed among 12 types. An obvious point of critique could be the idea that the set of equations I designed was close enough to a set of fundamental situations in relation to lower secondary school. For the sake of these studies, I chose this set of equations; however, I realise and know that such a set could be quite different in construction. One important factor was the choice of using the input field. The equations already implemented in MatematikFessor were all multiple choice, and I strongly feel that a lot has been learned from exploring the input field option instead.

In relation to the first part of the research question, I explored how the design principles might look if the tasks could detect students experiencing difficulties related to the equals sign. I conducted a somewhat theoretical thought experiment. The idea that students' actions (schemes) when solving tasks should reveal their thought process would only work to a certain extent when we can observe only their inputs in MatematikFessor. The application of the action, the four aspects of task design, led me towards the idea of designing tasks for digital environments that were of a diagnostic character and tied to the idea of ensuring that every aspect of the action that led the student to reach an answer could potentially be examined in the process. This led to the establishment of the dual scheme idea that can be applied to a specific learning goal, in this case the idea of substituting equal expressions in a task:

What number should be on the empty line?

$$3x \quad\quad = 11$$

$$3x + 4 = \_\_$$

The 'dual scheme' idea came about following the idea that this task could be solved in two obvious ways. The first is calculating the value of x and substituting it for the letter x in the second line's expression. The second is substituting the entire term of $3x$ with 11 in the send line's expression. We came up with the idea that the tasks could be presented in sequences where the calculation of the number x became increasingly difficult and where the substitution of the entire term would technically remain equally difficult. Afterwards, or in combination with a student either giving up or providing a wrong answer, follow-up questions based on the aspects of the scheme attempted to detect where in the action the student would experience difficulties.

This thought experiment is also a set of design ideas or principles aimed at structuring task design with detection or diagnostic purposes (for online learning environments), where only the answer is observed. The 'dual scheme idea' furthermore presents a learning opportunity, because in working with sequences of tasks where one solution method becomes increasingly more difficult, such a sequence might prove to let students discover the solution method that remains equally difficult.

6.2 Research Question 2

The second research question concerns the idea of diagnosing the cause of students experiencing difficulties solving equations. Where the first research question leaned into the idea of task design, the second focused more into the idea of analysing answers to such tasks.

What possibilities for the general diagnosis of students' difficulties related to equation solving can be established?

a. *What possibilities can be established based on a task found in the relevant literature?*
b. *What possibilities can be established based on the analysis of an exhaustive set of tasks involving equation solving.*

For this first part (a) of this research question, we worked with the famous task mentioned in Falkner et al. (1999): $8 + 4 = \_\_ + 5$. Additionally, I established the ideas of diagnostic potential and diagnostic value. Diagnostic potential is the task designer's a priori establishment of likely and possible student answers to a task that should indicate some behaviour leading to a diagnosis. Diagnostic value is the idea of evaluation the distribution of the actual answers a task receives, correct, interpretable and uninterpretable. The task from Falkner et al. (1999) presents a high diagnostic value. Close to no uninterpretable answer are given by students and the interpretable are all concerned with a particular interpretation of the equals sign.

In the first part of this second research question, I verified the high diagnostic value by replicating the task from Falkner et al. (1999) via MatematikFessor. The study (Paper D) confirmed that the diagnostic value of this task is very high, even when used in MatematikFessor. Even when answered by many students in an online learning environment, the task still received very few and very interpretable and meaningful answers.

The second part of this research question and perhaps the main attraction of this dissertation involved the analyses of the 892 equations designed under Research Question 1. In the analysis of these tasks, I implemented the coding system aimed at explaining or reasoning about the error(s) students made arriving at the solution given to the task in MatematikFessor. The coding system was applied to the five most popular answers for each of the 892 tasks. Thereafter, I conducted two separate analyses of the coded tasks:

1. Statistical overview of the representation of each code within each type of equation
2. Unsupervised learning method for establishing co-clusters of students and coded answers

In the statistical overview, we saw quite clearly the difference in errors students made relative to the structure or the type of equations they interacted with. Furthermore, I demonstrated some new errors, such as incorrectly rearranging terms and/or numbers in expression for them to make sense, not accepting negative numbers or confusing the solution with the additive inverse or presumably accidentally pressing the button next to the intended (or the actual solution) button on the keyboard when inputting the answer in MatematikFessor. Additionally, I confirmed many of the known errors students make when solving equations found in the literature over the last 45 years.

The unsupervised learning method of co-clustering the student × task matrix, with entries corresponding to the 94,368 participating students' coded answers to the 892 tasks, resulted in some interesting analyses that made it possible to observe the four interesting groups of students and their characteristics, as well as the almost 400 groups of tasks and their structure. Additionally, we recreated the co-clustering using a subset of the student × task matrix, only keeping the entries where a student had answered at least 100 of the tasks. These results yielded half as many task groups and a strengthened structure of student groups. Interestingly, these results yielded the opportunity to explore groupings of tasks (equations) independent of type, which performed highly similarly in terms of errors made in correspondence with each of the student groups. These results enable not only possibilities for diagnosing students as equation solvers in terms of the error types they make but also a strengthened possibility for task designers to evaluate their design and discuss the diagnostic value of the tasks in type-independent groupings.

As mentioned, a discussion about the finding under Hypothesis 2.1 in relation to diagnostic value is appropriate. Going through the types of equations I had designed one by one, led to I finding that can be considered as a task with negative diagnostic value. In the case of the equation $\frac{3}{x} + 2 = 5$ and the error type 'mv' (treating the entire fraction as the unknown), I found, that not only this reason could have led to the answer 3. In the same section I found indications that the idea of $\frac{a}{a} = a$, was applied. In the case of $\frac{3}{x} + 2 = 5$, both these ideas could have led to the answer 3. Because two recognised mathematical erroneous actions could have led to the same answer, the task is unreliable in terms of diagnostic value, meaning that the diagnostic value of this particular task is negative. This example underlines the importance of evaluating the diagnostic value and not only the diagnostic potential of tasks that we wish to base assessment on. In this case it is simply a matter of using different numbers in the equation in order to reach a task that does not present negative diagnostic value.

6.3 Research Question 3

The third research question was intended to cover the idea of establishing interventions that counteract specific difficulties identified in a diagnosis process. The ideas I presented in this dissertation lean into interventions from a design perspective. The first part of the research question is about designing an intervention that works with the interpretation and role of negative numbers in equations. The second part of the research question leans into the idea of establishing improved automated (formative) feedback for teachers working with equations in an online learning environment, such as MatematikFessor.

What possibilities for general interventions in relation to the concept of linear equations and equation solving can be established?

a. *What possibilities can be established based on a specific difficulty from the literature?*
b. *What possibilities can be established based on difficulties measured in an exhaustive set of tasks involving equation solving?*

For the first part (a) of this last research question, I explored the idea of establishing a situation or a space where students and teachers would be able to explore equations made abstract by negative numbers. The virtual environment we created gave students the opportunity to, in real time, use the classical balance model with the added feature of responding correctly to the 'negative weight' of terms with negative value. Our findings included, in addition to the design of the environments, a study of students' reactions to working in the environment with us as teachers/guides. The results showed positive prospects for the idea of being able to invert the weight of a term and restoring balance by placing the inverted term on the opposite balance pan. We found that working without mathematical operators as one does when solving equations using the balance model proved to be difficult in the attempt to transfer the knowledge to pen and paper exercises with mathematical operators separating the terms. The metaphor of the weight of collected items is that the sum of the individual weights is perhaps not so intuitive when working with pen and paper representations. We suggested implementing a small blackboard in the virtual environment that would serve as a live symbolic representation of the equation the user would work with live on the balance.

For the second part of the research question, I return to the set of equations, designed under the first research question and analysed under the second research question. I return to the overarching idea of enabling developers of online learning environments to create new and improved feedback for teachers based on the coding system and the analysis of tasks presented through the exploration of these research questions. The results showed that we could locate groups of tasks and expect only a certain number of different errors from students when solving these tasks. I argue that a feedback system based on developers of online learning environments applying codes to tasks through automated scoring would enable simple statistics to map out students as equation solvers. For example, this could be based on what types of equations students struggle with and issues related to their incorrect answers across multiple types of equations. I realise that what I am suggesting is not necessarily simple; however, much of the work has already been completed through this PhD project, which also offers an initial design for implementation that could be manufactured.

6.4 Reflections on the overall project and the motivation for it

The project set out to investigate data extracted from the online learning environment MatematikFessor. My expectations were that the data would suffice as a means for interpretation and analyses that would result in meaningful diagnoses that could lead to improved learning in the classroom. In the beginning, I thought that I would, after three years, have developed a digital tool for implementation in MatematikFessor that would improve students' learning experiences by personalising them based on the difficulties they struggle with.

I realised that developers claim to have been working on these adaptive tutoring systems for many years; however, I always expected them to be based on too simple structures and too few relevant assumptions about the didactical parameters. With didactical parameters I mean the general teaching material in Denmark is based on limited experience, and to a short extent on findings gathered through research.in mathematics education Perhaps research in mathematics education struggle or have yet, to prove its relevance for developers of teaching materials?

Without restarting the motivation sections, I was a good portion of the way into the project before realising that I wanted to address and possibly qualify teachers' formative work with the ideas and findings from the included papers in this project. I thought I would be able to finish such a digital tool that could be implemented into the feedback systems in MatematikFessor. Following the framework of detecting, diagnosing and intervening, the idea of working directly with the students to personalise their experience was perhaps easier to conceptualise than realise. However, I realised the interventions I could design and implement would possibly end up being too generic for what I imagined would actually be able to move the students forward. I imagined that such a tool would keep the students on the 'digital escalator' and reduce the teacher to, in the worst case, technical support.

The deviation from this idea of only working with students through MatematikFessor had me rethinking the idea of interventions, as I wanted to keep the idea of working with the framework from the Maths Counsellor Programme. The aspect of formative assessment became much more present, shifting to the idea of letting the data analyses inform teachers using MatematikFessor with their students. Much of the initial idea of detecting and diagnosing students' difficulties was still intact and developed into the six paper contributions and the *kappa*. I believe it was the right choice for me to address teachers instead of students. If I had the opportunity to continue the project, the next step would be to apply the results of analysing the codes in classrooms and conduct interviews with students to further verify the issues presented by the tasks. Importantly, I would conduct interviews with teachers to gather information about how such automated feedback would influence their teaching and how such feedback would be received. An obvious limitation of this project is the natural extension of the last sentence. All my findings have not been tested or discussed with teachers using MatematikFessor. I am hopeful that such a world exists where teachers would benefit positively from an improved type of feedback from MatematikFessor and that their teaching would flourish and be more fun and relevant to their students. I would encourage other researchers to study the use of statistical tools in online learning environments at the classroom level. I believe that it is important to further develop knowledge on how teachers identify formative values through these statistical tools and offer extended insight into students' thoughts on how these statistical tools affect their classroom culture.

As mentioned in the section about papers and collaborations, I cannot place enough emphasis on how collaborations with experts from within the field of mathematics education and perhaps even more so with experts from outside the field have been a priority of mine. Strengthening the possibilities for creating advanced data analyses and creating new and innovative grounds for teaching linear equations were of utmost importance to me. As an industrial PhD student, I participated in seminars and workshops with other industrial PhD students from a variety of fields far from mathematics education. Meeting researchers from research fields such as pharmacy, engineering and construction greatly inspired me to reach outside the sometimes seemingly closed field of mathematics education for perspectives and ideas that would benefit my project.

6.5 Discussion of DE as a structure for a *kappa*

In this final chapter, I wish to reflect on the process of applying the structure of DE as a method for writing a *kappa*. I remind the reader that while establishing the research questions, I posed an additional question referring to writing a *kappa*:

> *In what sense can DE serve as an umbrella for structuring a kappa as part of a PhD dissertation?*

Returning to the structure of DE as a research method, I remind the reader of the important point that the structure was applied in retrospect. What this effectively means for the reading of this dissertation is that the chronological aspect of DE from phases one through four was implemented only for the sake of reading this dissertation. The papers feeding into the different phases of the DE initially only satisfied the overall idea presented through the research questions in relation to the Maths Counsellor Programme's framework of detection, diagnosis and intervention. The four phases of DE, specifically the preliminary analyses, a priori analysis, realisations and a posteriori analysis, were applied to better guide the reader through the project and its contributions.

With that said, I genuinely believe that the application of the phases and the structure of DE helped me organise the paper contributions into a dissertation that is better or more cohesive. In the process of writing this *kappa*, I forced myself to evaluate each and every one of the phases to not only guide the reader but also to achieve a more valid and trustworthy industrial PhD project dissertation.

I am far from a DE expert, and I realise that the method perhaps is mainly reserved for didactical design and for the design of teaching. However, I propose that this method is also valuable as a structure for presenting PhD dissertations that are paper collections. In all its simplicity, the structure of DE makes sense. The idea of learning about or analysing an environment before establishing hypotheses about how this environment would react if exposed to some innovation makes sense. After some empirical realisations based on applying the innovation, one would also naturally attempt to confirm or reject the posed hypotheses. To a large extent, this is what the scholarly method is and should be.

The structure of DE takes this intuitive framework and applies it to the science of teaching in all its aspects. Through DE, we observed nuances about the three areas of the environment the preliminary analyses should revolve around. Furthermore, we are advised to include a historical perspective in these preliminary analyses. Thus, I found the structure of DE extremely helpful and meaningful, not only as a research method but mostly as a guide for me to structure this dissertation.

However, as mentioned the project was not planned with DE in mind. DE was only used for structuring and writing this *kappa*. I attempted to establish the hypotheses during the writing of the conception and a priori analysis, which I was actually working with throughout the project. These hypotheses would just have been structural ideas that would possibly not have been for the reader to review had the DE structure not been applied. To be fair, I added hypotheses to Phase Two while writing up this dissertation. Therefore, the hypotheses included might not be all the hypotheses I worked with during the project. However, these are the hypotheses that I deemed relevant for the reader to be familiar with to be able to read this dissertation and to ensure the relevance of the included paper contributions.

Applying the DE structure retrospectively requires honesty. The hypotheses stated in the a priori analysis must be honest and relevant to the project. To a large extent, the relevance is self-explanatory if the framework of DE is meant to help guide the reader through a dissertation. To that end, DE serves a clear purpose of structuring the relevant aspects of a research project in a cohesive manner. Regarding honesty, the author should, while ensuring readability, attempt to tell the actual story of how the research came about. These thoughts were what went into establishing the hypotheses for this dissertation. These are not to be confused with the research questions presented in the paper contributions. The hypotheses established in the a priori analysis phase of this *kappa* is my idea of how I wanted to explore the research *problématique* established in the introduction.

I took some level of freedom when applying the structure of DE. In Phase Four, I realised that the DE method requires the researcher to engage in discussion before confirming or rejecting the established hypotheses. I took the liberty to discuss and evaluate them on three levels. First, I discussed the convergences and divergences of the hypotheses as prescribed. Second, I discussed what I had learned from the preliminary analyses and how this affected my findings from exploring the hypotheses and writing the papers included in this dissertation. Third, I returned in the concluding remarks to the posed research questions as part of the research *problématique* and the connection to the overall idea of following the framework prescribed in the Math Counsellor Programme.

6.6 For the future

These is a lot to be learned about task design for online learning environments, in particular in relation to the effect task design has on the feedback system. If I was to continue this research project or start it again I think that a more comprehensive analysis of teachers' use of the statistical tools in MatematikFessor would be appropriate. Many of the choices made during this PhD project, I was only able to make late in the process. I imagine that this holds true for many PhD projects. However, looking back on the process, I wish that I had been wiser in the beginning. The structure of DE came in late in the process of this PhD project. However, ideas of the design process of DE were implemented already. The preliminary analyses make sense in the way that one should be familiar with the environment and its traditions and its stakeholders. Thereafter, one construct hypotheses about how an innovation can or will affect this environment. Realisations exploring the hypotheses are made, before the hypotheses could be confirmed or rejected. I imagine that such generalisation of engineering or design holds true in many contexts. Returning to the research field of mathematics education, I believe that more projects could benefit from comprehensive preliminary analyses.

I believe that research and development share a joint responsibility for enabling good teaching in Denmark. If it bears any truth that feedback systems in digital environments shape teachers' perception of how students' learning and understanding is signified through their proportion of correct answers, I believe that both teachers and developers are missing the point. Feedback systems for teachers should be relevant and I think there is much more to be learned about what such relevance is and how to improve it.

References

Aguilar, M. S. (2020). Replication studies in mathematics education: What kind of questions would be productive to explore? *International Journal of Science and Mathematics Education*, *18*(1), 37–50. https://doi.org/10.1007/s10763-020-10069-7

Artigue, M. (2009). Didactical design in mathematics education. In C. Winslow (Ed.), *Nordic research in mathematics education* (pp. 5–16). Brill Sense.

Artigue, M. (2014). Didactic Engineering in Mathematics Education. In S. Lerman (Ed.), *Encyclopedia of Mathematics Education* (pp. 159–162). Springer Netherlands. https://doi.org/10.1007/978-94-007-4978-8_44

Artigue, M. (2015). Perspectives on Design Research: The Case of Didactical Engineering. In A. Bikner-Ahsbahs, C. Knipping, & N. Presmeg (Eds.), *Approaches to Qualitative Research in Mathematics Education: Examples of Methodology and Methods* (pp. 467–496). Springer Netherlands. https://doi.org/10.1007/978-94-017-9181-6_17

Artigue, M., Haspekian, M., & Corblin-Lenfant, A. (2014). Introduction to the theory of didactical situations (TDS). In *Networking of theories as a research practice in mathematics education* (pp. 47–65). Springer.

Bachelard, G. (1938). *La formation de l'esprit scientifique*. Vrin.

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

Bodin, A. (1993). What does to assess mean? The case of assessing mathematical knowledge. In *Investigations into assessment in mathematics education* (pp. 113–141). Springer.

Boell, S. K., & Cecez-Kecmanovic, D. (2014). A Hermeneutic Approach for Conducting Literature Reviews and Literature Searches. *Communications of the Association for Information Systems*, *34*. https://doi.org/10.17705/1CAIS.03412

Bokhove, C. (2017). Supporting Variation in Task Design Through the Use of Technology. In *Digital Technologies in Designing Mathematics Education Tasks* (pp. 239–257). Springer.

Brousseau, G. (1997). *Theory of didactical situations in mathematics (N. Balacheff, M. Cooper, R. Sutherland, & V. Warfield, Eds. & Trans.)*.

Brousseau, G. (2006). *Theory of didactical situations in mathematics: Didactique des mathématiques, 1970–1990* (Vol. 19). Springer. https://doi.org/10.1007/0-306-47211-2

Cavalcanti, A. P., Barbosa, A., Carvalho, R., Freitas, F., Tsai, Y.-S., Gašević, D., & Mello, R. F. (2021). Automatic feedback in online learning environments: A systematic literature review. *Computers and Education: Artificial Intelligence*, *2*, 100027. https://doi.org/10.1016/j.caeai.2021.100027

Collis, K. F. (1978). Operational thinking in elementary mathematics. *Cognitive Development: Research Based on a Neo-Piagetian Approach*, 221–248.

Davis, R. B. (1975). Cognitive Processes Involved in Solving Simple Algebraic Equations. *Journal of Children's Mathematical Behaviour*, *1*, 7–35.

Dawson, P., Henderson, M., Ryan, T., Mahoney, P., Boud, D., Phillips, M., & Molloy, E. (2018). Technology and feedback design. *Learning, Design, and Technology*. https://doi.org/10.1007/978-3-319-17727-4_124-1

Falkner, K. P., Levi, L., & Carpenter, T. P. (1999). Children's Understanding of Equality: A Foundation for Algebra. *Teaching Children Mathematics*, *6*(4), 232–236. https://doi.org/10.5951/TCM.6.4.0232

Filloy, E., & Rojano, T. (1989). Solving Equations: The Transition from Arithmetic to Algebra. *For the Learning of Mathematics*, *9*(2), 19–25. https://www.jstor.org/stable/40247950

Gallardo, A. (2002). The extension of the natural-number domain to the integers in the transition from arithmetic to algebra. *Educational Studies in Mathematics*, *49*(2), 171–192. https://www.jstor.org/stable/3483074

Herlau, T., Smith, M. N., & Mørup, M. (2022). *Introduction to Machine Learning and Data Mining* (1.0). DTU.

Herscovics, N., & Kieran, C. (1980). Constructing meaning for the concept of equation. *NCTM*, *78*(8), 9–9. https://www.jstor.org/stable/27962179

Høyrup, J. (1998). *Algebra på lertavler*. Matematiklærerforeningen.

Jankvist, U. T., & Niss, M. (2015). A framework for designing a research-based "maths counsellor" teacher programme. *Educational Studies in Mathematics*, *90*(3), 259–284. https://doi.org/10.1007/s10649-015-9629-8

Jankvist, U. T., & Niss, M. (2017). The notion and role of" detection tests" in the Danish upper secondary" maths counsellor" programme. In T. Dooley & G. Gueudet (Eds.), *Prooceedings of the Tenth Congress of the European Society for Reseach in Mathematics Education* (pp. 3825–3832). Institute of Education, Dublin City University and ERME.

Katz, V. J. (2009). *A History of Mathematics* (3rd ed., Vol. 2009). Addison-Wesley.

Kieran, C. (1981). Concepts Associated with the Equality Symbol. *Educational Studies in Mathematics*, *12*(3), 317–326. https://doi.org/10.1007/BF00311062

Kieran, C. (1985). The equation-solving errors of novice and intermidiate algebra students. In L. Streefland (Ed.), *Proceedings of the Annual Conference of the International Group for the Psychology of Mathematics Education* (pp. 141–146).

Kieran, C. (2007). Learning and teaching of algebra at the middle school through college levels: Building meaning for symbols and their manipulation. In F. K. Lester (Ed.), *Second Handbook of Research on Mathematics Teaching and Learning* (Vol. 2, pp. 707–762). NCTM-Information Age Publishing.

Küchemann, D. (1981). Algebra. In K. Hart (Ed.), *Children's Understanding of Mathematics: 11-16* (pp. 102–119). Murray.

Linsell, C. (2008). Solving equations: Students' algebraic thinking. *Findings from the New Zealand Secondary Numeracy Project 2007*, 39–44.

Linsell, C. (2009a). A Hierarchy of Strategies for Solving Linear Equations. In R. Hunter, B. Bicknell, & T. Burgess (Eds.), *Crossing divides: Proceedings of the 32nd annual conference of the Mathematics Education Research Group of Australasia* (Vol. 1, pp. 331–338). MERGA.

Linsell, C. (2009b). Students' knowledge and strategies for solving equations. *Findings from the New Zealand Secondary Numeracy Project 2008*, 29–43.

MacGregor, M., & Stacey, K. (1993). Cognitive models underlying students' formulation of simple linear equations. *Journal for Research in Mathematics Education*, *24*(3), 217–232. https://doi.org/10.2307/749345

Marton, F. (2015). *Necessary conditions of learning*. Routledge.

Matthews, P., Rittle-Johnson, B., McEldoon, K., & Taylor, R. (2012). Measure for Measure: What Combining Diverse Measures Reveals about Children's Understanding of the Equal Sign as an Indicator of Mathematical Equality. *Journal for Research in Mathematics Education*, *43*(3), 316–316. https://doi.org/10.5951/jresematheduc.43.3.0316

Ministry of Children and Education. (2019a). *Matematik Fælles Mål (Common goals for mathematics)*. Danish Ministry of Children and Education. https://emu.dk/sites/default/files/2020-09/GSK_F%C3%A6llesM%C3%A5l_Matematik.pdf

Ministry of Children and Education. (2019b). *Matematik Læseplan (Lesson Plan for Mathematics)*. Danish Ministry of Children and Education. https://emu.dk/sites/default/files/2020-09/GSK_L%C3%A6seplan_Matematik.pdf

Niss, M., & Højgaard, T. (2019). Mathematical competencies revisited. *Educational Studies in Mathematics*, *102*(1), 9–28. https://doi.org/10.1007/s10649-019-09903-9

Otten, M., Van den Heuvel-Panhuizen, M., & Veldhuis, M. (2019). The balance model for teaching linear equations: A systematic literature review. *International Journal of STEM Education*, *6*(1), 30–30. https://doi.org/10.1186/s40594-019-0183-2

Parker, D. H. (1929). On the Notion of Value. *The Philosophical Review*, *38*(4), 303–325. JSTOR. https://doi.org/10.2307/2180021

Pirie, S. E. B., & Martin, L. (1997). The equation, the whole equation and nothing but the equation! One approach to the teaching of linear equations. *Educational Studies in Mathematics*, *34*(2), 159–181.

Prediger, S. (2010). How to develop mathematics-for-teaching and for understanding: The case of meanings of the equal sign. *Journal of Mathematics Teacher Education*, *13*(1), 73–93. https://doi.org/10.1007/s10857-009-9119-y

Rezat, S. (2021). How automated feedback from a digital mathematics textbook affects primary students' conceptual development: Two case studies. *ZDM–Mathematics Education*, *53*(6), 1433–1445. https://doi.org/10.1007/s11858-021-01263-0

Rhine, S., Harrington, R., & Starr, C. (2018). *How Students Think When Doing Algebra*. Information Age Publishing, Incorporated.

Sfard, A. (1991). On the dual nature of mathematical conceptions: Reflections on processes and objects as different sides of the same coin. *Educational Studies in Mathematics*, *22*(1), 1–36. https://doi.org/10.1007/BF00302715

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, *78*(1), 153–189. https://doi.org/10.3102/0034654307313795

Smith III, J. P., diSessa, A. A., & Roschelle, J. (1994). Misconceptions Reconceived: A Constructivist Analysis of Knowledge in Transition. *Journal of the Learning Sciences*, *3*(2), 115–163. https://doi.org/10.1207/s15327809jls0302_1

Spitzer, M. W. H., & Musslick, S. (2021). Academic performance of K-12 students in an online-learning environment for mathematics increased during the shutdown of schools in wake of the COVID-19 pandemic. *PloS One*, *16*(8), e0255629. https://doi.org/10.1371/journal.pone.0255629

Steenbergen-Hu, S., & Cooper, H. (2013). A meta-analysis of the effectiveness of intelligent tutoring systems on K–12 students' mathematical learning. *Journal of Educational Psychology*, *105*(4), 970. https://doi.org/10.1037/a0032447

Sun, J., Przybylski, R., & Johnson, B. J. (2016). A review of research on teachers' use of student data: From the perspective of school leadership. *Educational Assessment, Evaluation and Accountability*, *28*(1), 5–33. https://doi.org/10.1007/s11092-016-9238-9

Tall, D., & Vinner, S. (1981). Concept image and concept definition in mathematics with particular reference to limits and continuity. *Educational Studies in Mathematics*, *12*(2), 151–169. https://doi.org/10.1007/BF00305619

Trouche, L., Drijvers, P., Gueudet, G., & Sacristán, A. I. (2013). Technology-Driven Developments and Policy Implications for Mathematics Education. In M. A. (Ken) Clements, A. J. Bishop, C. Keitel, J.

Kilpatrick, & F. K. S. Leung (Eds.), *Third International Handbook of Mathematics Education* (pp. 753–789). Springer New York. https://doi.org/10.1007/978-1-4614-4684-2_24

Utterberg Modén, M. (2021). *Teaching with Digital Mathematics Textbooks-Activity Theoretical Studies of Data-Driven Technology in Classroom Practices*.

Vergnaud, G. (1988). Multiplicative structures. *Number Concepts and Operations in the Middle Grades*, 141–161.

Vergnaud, G. (1998). A comprehensive theory of representation for mathematics education. *The Journal of Mathematical Behavior*, *17*(2), 167–181. https://doi.org/10.1016/S0364-0213(99)80057-3

Vergnaud, G. (2009). The theory of conceptual fields. *Human Development*, *52*(2), 83–94. https://doi.org/DOI: 10.1159/000202727

Vlassis, J. (2002). The balance model: Hindrance or support for the solving of linear equations with one unknown. *Educational Studies in Mathematics*, *49*(3), 341–359. https://doi.org/10.1023/A:1020229023965

Watson, A., & Mason, J. (2006). Seeing an exercise as a single mathematical object: Using variation to structure sense-making. *Mathematical Thinking and Learning*, *8*(2), 91–111. https://doi.org/10.1207/s15327833mtl0802_1

Wiliam, D., & Thompson, M. (2007). Integrating assessment with learning: What will it take to make it work? In *The future of assessment* (1st ed., pp. 53–82). Routledge. https://doi.org/10.4324/9781315086545

Paper A

Elkjær, M., & Hansen, C. (2020). Clustering student errors. In B. Barzel, R. Bebernik, L. Göbel, M. Pohl, H. Ruchniewicz, F. Schacht, & D. Thurm (Eds.), *Proceedings of the 14th International Conference on Technology in Mathematics Teaching – ICTMT 14, Essen, Germany,* 165–172.

CLUSTERING STUDENT ERRORS

Morten Elkjær[1] & Christian Hansen[2]

[1]*Danish School of Education, Aarhus University, Emdrup, Denmark; me@edu.au.dk*

[2]*University of Copenhagen, Copenhagen, Denmark; christianh@edulab.dk*

This paper gives an in-depth description of the research design in the pursuit of clustering of students' performance when solving different types of linear equations. The student performances are clustered using data from 457,185 answers to equation tasks, made by 37,585 students, distributed across 3,438 unique linear equations in a digital learning environment. The tasks consist of different categories of linear equations. The clustering analysis contributes to the development of an online tool to provide the teachers with easy accessible formative assessment. At this point, the attempt to cluster the students' performance have not yet been successful, meaning that no clusters are found. Instead, a description of how the pursuit of these clusters will continue is presented alongside the research design.

Keywords: Students' difficulties, linear equations, unsupervised learning, clustering, formative assessment.

INTRODUCTION

This paper presents the research design for utilizing a large amount of data to assess and accommodate students' mathematical difficulties when they are working with linear equations in Danish lower secondary school, more specifically the 6th and 7th grade (12-14 years old). Overall, this paper describes a research design that is based around the clustering of students working with linear equations in an online learning environment. An initial attempt was made, that unfortunately did not bear fruit. The project stems from a collaboration between two doctoral candidates associated with two independent industrial PhD projects. The authors collaborate across academic institutions with the common goal of generating knowledge about whether or not standard digital mathematical tasks on an online platform can serve as a non-disruptive diagnostic tool to generate easily accessible formative assessment for the teachers in Danish lower secondary school. The ideas and the design described are research work in progress. Therefore, only initial results and anticipated outcomes will be presented alongside the design.

The Danish private company Edulab develops and maintains an online mathematical learning platform for the Danish K-10 schools. In Denmark, 75% of the K-10 schools subscribe to Edulab's learning platform, which effectively means that 600,000 students have access to these digital learning materials. During the last 12 years, Edulab has developed the online mathematical learning platform, called *matematikfessor.dk*. The platform performs primarily as a teacher-driven supplement to a learning material, but the students also have access in order to explore the mathematical content on their own. Every day Danish school students collectively give answers to 1.5 million tasks on the online learning platform. This creates a unique opportunity to implement didactical research results regarding students' difficulties with equations directly into practice. The two industrial PhD projects aim to provide Edulab with a research-based tool to reveal and capture students' mathematical difficulties when working with linear equations in order to provide teachers with valuable information hereof. The assumption is that standard digital mathematical tasks (already implemented on the platform) can serve as a substitute for a diagnostic test, based on the idea that a large enough amount of data together with research-based diagnostic verification can generate significant information on students with difficulties when working with linear equations. This leads to the following overall research aim:

How can an online diagnostic tool for lower secondary school be designed, utilizing existing research findings on mathematical difficulties when working with equations in order to provide the teachers with significant dynamic formative assessment?

In order to approach the answer to the above question, we have chosen to explore possibilities based on the vast amount of data that Edulab has collected so far from students answering textbook like standard tasks, involving equations, in their online environment. Therefore, we have posed the following research question that will be the main research question for this paper:

To what extent can the existing categorization of standard textbook linear equations serve as a mean for generating clusters of students with a large amount of data?

By existing categorization, we mean the levels of distinctions we are able to make based on the tasks already implemented on the planform.

THEORETICAL BACKGROUND

Terms like misconceptions, alternative conceptions, or errors have been used in the mathematics education literature in the past describing children's beliefs, conceptions and problem solving strategies (J. L. Booth, McGinn, Barbieri, & Young, 2017; L. Booth, 1984; Linsell, 2009). When we mention children's difficulties working with equations in this particular context, the main focus is to describe common misinterpretations or common conceptual limitations that children have about linear equations, for example, the role of the equal sign, the role of the literal symbols or the ability to apply different problem solving strategies.

From reviewing the literature, it is concluded that the mathematics educations community knows a great deal about children's conceptions and errors working with the elements of algebra (Rhine, Harrington, & Starr, 2018). Linear equations have not received the same attention, but some studies have dealt with the strategic aspect and the concrete errors children make when working with equation solving (Herscovics & Linchevski, 1994; Kieran, 1985, 1992; Linsell, 2009). Linsell (2009) presents tests and interviews with the purpose of uncovering the difficulties children have with solving equations as well as the origin of these difficulties. Together with the strategic difficulties that arise when students have to solve equations, there are of course also difficulties with the individual sub-concepts of linear equations. The interpretation of the equal sign becomes a natural conceptual centrepiece when addressing the overall concept of equations (Kieran, 1981). Alongside the equal sign comes the algebraic elements of literal symbols, numbers and operators (L. Booth, 1984; Küchemann, 1981). When talking about the concept of equations, one also has to take more intangible aspects such as truth-value and solution into consideration.

Following categorization of the types of linear equations is adapted from Vlassis (2002) based on abstraction level and the partition presented in Filloy and Rojano (1989).

- Concrete arithmetical equations: Arithmetical equations that consist only of natural numbers and only include a single occurrence of the unknown. (e.g. $ax + b = c, \quad a, b, c \in N_0$)

- Abstract arithmetical equations: Equations with the unknown in one member, which require certain algebraic manipulations, because of the presence of negative integers or several occurrences of the unknown. ($a_1x \pm \cdots \pm a_n \pm b_1x \pm \cdots \pm b_nx\ \pm c,\quad a_1 \ldots a_n, b_1 \ldots b_n, c\ \in Z,\ \ n \in N$)

- Algebraic equations: Equations similar to the abstract arithmetical equations, but with the exception of occurrences of the unknown on both sides. ($ax \pm b = cx \pm d,\quad a, b, c, d\ \in Z$)

Linsell (2007) lets us divide this categorization, into equation types based on the number of steps it will take to solve the equations by normal transformations. Within the concrete arithmetical equations, we find the one-step equations, divided into containing small and large numbers. The abstract arithmetical equations contain (in this context) two- and three-step equations. It is also possible to have two-step equations within the algebraic equations, but for the most part in this context, they will require further transformational steps. The categorization of the equations used for analysis, taken from the online platform, is presented in table 1.

Label	**Form**	**Steps**	**Number of tasks**
1	$ax + b = c,\qquad x, a, c \in \{1,2, \ldots, 9\}, b \in \{0,1, \ldots, 9\}$	One/Two-step	106
2	$ax + b = c,\qquad x, a, c \in N\ b \in N_0$	One/Two-step	812
3	$ax + b = c,\qquad x, c \in Z\ b \in N_0, a \in N$	One/Two-step	501
4	$ax + b = cx + d,\qquad a, b, c, d, x\ \in N$	Two/Three-step	1478
5	$ax + b = cx + d,\qquad a, c\ \in N\ x, b, d\ \in Z$	Two/Three-step	519

Table 1 - Categorization of linear equations on the online platform

The reason for having this categorization of the linear equation is that the online platform currently only give us the opportunity to make the above distinction at this point. Notice that equations labelled 1, 2 and 3 are arithmetical, but are only considered abstract if e.g. c is a negative number. The equations labelled 4 and 5 are considered algebraic equations.

METHODOLOGICAL CONSIDERATIONS

Design research (Barab & Squire, 2004) will pave the way for the knowledge generated through iterations of the work with the initial clustering of students performance solving different types of equations and the selection of the standard textbook tasks, about equations, implemented on the online learning platform.

In order to answer the main research question for this paper, the design goal is the analysis of student interaction with the chosen items on the platform, fitting the categorization based on the literature. The categorization of the equations mentioned in the theoretical background is to be utilized in order to analyse the clusters when found. In the following section, we describe methodological considerations as well as the process of and theory behind the approach for the clustering of the different types of students based on the answers given to the equations presented in the theoretical background and the specific errors the student make while working with these equations.

CATEGORIZING AND CLUSTERING OF ERRONEOUS ANSWERS AT EDULAB

This section contains our current approach to the categorizing and clustering of the errors the students make when using the portal. Before the approach can be described in detail, it is necessary to shortly establish how the current system at Edulab works.

The primary activity students engage in at Edulab's online learning platform is answering tasks. The tasks can be either multiple-choice or writing an answer in an input field. The multiple-choice tasks are structured such that each task always has five possible answers, while the free form input fields can only contain a number. Each task is associated with a lesson e.g. "Equations with a single unknown, plus minus", which is the lesson introducing them to equations with only one unknown, containing only simple plus or minus operations. A lesson is a categorized element consisting of a video-clip introducing the content and related tasks. Each lesson is further associated to a topic, e.g. "Equations with unknowns" which is more a general element than the lesson.

Edulab's platform contains a collection of over 1 million tasks. Each task is therefore not "handmade", and there is no meta-information related to the task on what different erroneous answers can indicate of possible underlying mathematical difficulty related to equations. Despite the lack of meta-information, we hypothesize that with a large number of erroneous answered tasks from the users on the portal, we can infer what the erroneous answers might indicate of student difficulties.

Distribution of erroneous answers

Before we elaborate on how we intend to use the erroneous answers, we first investigate how the erroneous answers are distributed for each task to ensure that not all erroneous answers are equally likely to be chosen by the students, as this will require us to infer the meaning of a very large number of erroneous answers. If some erroneous answers are very unlikely, we will not try to utilize these, as the information for these are very rare.

We will focus on the distribution of erroneous answers for a lesson related to equations, which only contain multiple-choice tasks. We use log data collected for the school year 2018-2019 for students attending 6^{th} and 7^{th} grade. The log contains 457,185 answers to tasks, provided by 37,585 students, in the lesson containing the equation tasks mentioned in table 1, distributed across 3,416 tasks. We remove all tasks that received less than 150 answers leaving us 197 tasks and 379,315 answers. To investigate the distribution of erroneous answers we apply the following for each task:

1. Compute the number of times each answer is answered by the students for each task
2. Normalize it to a probability distribution
3. Sort the probabilities from most probable to least probable (On all answers the most probable answer is the correct answer)

We are left with a probability distribution for each task, which we want to cluster, to see how the answers of the students vary across tasks. To do the clustering we use affinity propagation (Frey & Dueck, 2007), which finds candidate "exemplars" of the data points, which can be used as a general example for a cluster. We use standard parameters with damping being 0.5, and the preference is the median of the affinities. For the affinity measure, we use the well-known Jensen-Shannon divergence, as the data points we wish to cluster are probability distributions. Doing this we find 3 clusters:

1. Cluster 1 are tasks that primarily are always answered correctly, with very few errors.
2. Cluster 2 are tasks that have 1 particular wrong answer, which receives most of the erroneous answers, when a student chooses an incorrect answer.
3. Cluster 3 are tasks that have 2-3 wrong answers, which receive most of all the erroneous answers, when a student chooses an incorrect answer.

Cluster 1 have 88 of the tasks, cluster 2 have 39 and cluster 3 have 70. Our assumption that not all erroneous answers are equally likely is thereby supported.

We have done a similar analysis on a collection of other lessons, which also have input field answers, which shows that the answers from input fields also are focused on few very likely erroneous answers, and a large tail distribution of almost random answers.

Co-occurrence of erroneous answers

In the previous section we established that not all erroneous answers are equally likely, therefore the erroneous answers are focused on a smaller subset of possible erroneous answers. In the following, we establish how the erroneous answers are going to be utilized.

Each erroneous answer to a task gets a unique ID, if the erroneous answer has received more than some threshold, ***P***, of the answers for the given task. All erroneous answers, which occur often, will therefore have a unique ID, while we ignore the erroneous answers, which receive little attention. The reason for this is twofold, if all erroneous answers get an ID, the space of erroneous answers will be very large, and it is thus difficult to infer the meaning of each erroneous answer as they occur rarely.

Doing this, each student will have a sequence of erroneous answers which have been given an ID, $e_1, e_2, \dots, e_k$, ordered by when the student gave the answer. Based on this sequence we can now construct a co-occurrence matrix, which is a MxM where M is the number of erroneous answer IDs. Each row and column correspond to an erroneous answer, and the entry at the i'th row, j'th column is a counter of how often the i'th erroneous answer occur together with the j'th erroneous answer in a student sequence. What it means for two IDs to occur together is a matter of choice, e.g. if there is a sequence of student errors over a long time horizon, then the definition of two erroneous answers to occur together can depend on the time between the two erroneous answers. By doing this, only erroneous answers, which occur closely together in time, are considered a pair. On the other hand, all IDs for the same student can also be considered as occurring together if only a small number of lessons are considered, or the time horizon is small.

We use data for a single lesson (the same as in the previous section) which focuses on equations to construct the co-occurrence matrix, where we set the threshold ***P***=5%. We only include students who have made more than three mistakes on the lesson. This is early work, and for the final work, we wish to include a long series of lessons related to equations, but these are currently being deployed, and the data needs to be collected.

For this project we have no labels and are therefore limited to unsupervised learning (Hastie, Tibshirani, & Friedman, 2009), which is the paradigm of machine learning of learning some underlying structure of the data. As an initial experiment, we wanted to investigate if there was any clusters in the co-occurrence matrix, as this would indicate the existence of errors that mostly occurred together.

To explore the existence of clusters, we employed t-SNE (Maaten & Hinton, 2008), which is a powerful non-linear clustering technique, which tries to find a mapping for each data points to a new space, such that elements which are close in the original space, are also closed in the mapped space.

This exploration did not reveal any clusters, and we therefore did not manage to find any errors that occur primarily together.

When we get the full dataset, where students interact with equations over a much larger variety of lessons, we will repeat the following exploration. This is further elaborated on in the final section of the paper. In addition, the co-occurrence matrix allows for embedding based strategies of the errors. An embedding of an error would be some function $F(e_i)$ which maps the error to some real space of dimension N. The embedding would be such that errors who occur often together will be more similar than errors than does not occur together. This can be done using the algorithm Glove (Pennington, Socher, & Manning, 2014) which is used widely for finding embeddings for words, and work directly on a co-occurrence structure like ours.

CONCLUSION AND FUTURE DESIGN

The main focus of this paper was to answer the question; to what extent a categorization of standard textbook linear equations could serve as a mean for generating clusters of students with a large amount of data. Unfortunately, the clusters are not yet found. In the following section, we elaborate on the next step of the research design in order to accomplish finding the clusters.

To link the content of this paper to the overall research question, the following section will describe the future possibilities for finding clusters. With the right strategy, it will be possible to push these new tasks to the users of the platform, both for the teachers to assign to their students and for the students to explore on their own. In table 2 is presented a new possible categorization of the content on the platform that involves solving linear equations. This content already exists but have not yet been put to full use, meaning that the amount of answers given to these tasks are yet too sparse.

Label	Form		Type	Steps	Number of tasks
A	$x + a = b,$	$a \in N, b \in Z\backslash\{0\}$	Arithmetical	One-step	910
B	$ax = b,$	$a, b \in N$	Arithmetical	One-step	467
C	$x - a = b,$	$a \in N, b \in Z\backslash\{0\}$	Arithmetical	One-step	432
D	$a - x = b,$	$a \in N, b \in Z\backslash\{0\}$	Arithmetical	One-step	493
E	$ax + b = c,$	$a, b, c \in N$	Arithmetical	Two-step	431
F	$\frac{x}{a} = b,$	$a \in N, b \in Z\backslash\{0\}$	Abstract Arith.	One-step	426
G	$x \cdot \frac{a}{b} = c,$	$a, c \in Z\backslash\{0\}, b \in N$	Abstract Arith.	One-/two-step	587
H	$\frac{a}{x+b} = c,$	$a, c \in Z\backslash\{0\}, b \in N$	Abstract Arith.	Two-/three-step	755
I	$ax + b = cx + d,$	$a, c \in N\ b, d \in Z\backslash\{0\}$	Algebraic	Two-/three-step	DBT
J	$ax + b = cx + d,$	$a, c \in Z\ b, d \in Z\backslash\{0\}$	Algebraic	Two-/three-step	DBT

Table 2 – Future categorization of linear equations on the online platform

The future goal is to use the same clustering approach but with the new, more refined variety of linear equations. Furthermore, every type of equation comes in five different difficulty levels. This gives us another opportunity for selection and distinction. This means that instead of just having the 10 categories presented in table 2 we possibly have 50 levels of distinction instead of the 5 we had for the first iteration of the clustering attempt. We believe that having 50 levels of distinction between the different types of linear equations will be a step in the right direction in order to achieve the student clustering. As mentioned in the methodological considerations having this categorization of the types of linear equations will serve as a mean to analyse and interpret the clusters when found.

FINAL REMARKS

The overall research aim of the project is to provide Edulab with a tool to support teachers in their teaching. The utilization of the vast amount of data Edulab are able to collect shall pave the way for the development of this assessment tool. The project's vision is that Danish mathematics teachers, based on an easy accessible formative assessment source, can be given a unique opportunity to organize, plan and complete their teaching (Palm, Andersson, Boström, & Vingsle, 2017). The future goal is to set up a continuous categorization of all students using Edulab's platform in the 7th grade in order to "catch" students with conceptual or strategical difficulties when working with linear equations on the platform.

Inspired by the work presented in Linsell (2009), the goal is to develop a series of online questions in order to uncover the strategies the students use or the transformational difficulties the students experience, when working with the different types of linear equations. Together with the categorization of students, the second goal is to provide teachers with information of students' strategies as well as typical errors their students make when solving different types of equations. This information should be in the form of auto-generated 'formative assessment reports' for each student that the tool has identified and verified as having conceptual difficulties working with linear equations, describing also the student's behaviour and the difficulties present based on research findings on the identified difficulties. The formative assessment report should contain procedures and guidelines for how the teacher can approach remediation of the student's identified difficulties.

An alternate way to proceed could be to provide the student suggestions for teaching materials on the online learning platform for further learning on their own. The suggestions could be video lectures addressing their concrete subject in which they have difficulties.

REFERENCES

Barab, S., & Squire, K. (2004). Design-based research: Putting a stake in the ground. *The Journal of the Learning Sciences, 13*(1), 1–14.

Booth, J. L., McGinn, K. M., Barbieri, C., & Young, L. K. (2017). Misconceptions and learning algebra. In S. Stewart (Ed.), *And the rest is just algebra* (pp. 63–78). Cham: Springer.

Booth, L. (1984). *Algebra: Children's strategies and errors. A report of the strategies and errors in secondary mathematics project.* Windsor, Birkshire: NFER-NELSON Publishing company Ltd.

Filloy, E., & Rojano, T. (1989). Solving Equations: The Transition from Arithmetic to Algebra. *For the Learning of Mathematics, 9*(2), 19–25.

Frey, B. J., & Dueck, D. (2007). Clustering by passing messages between data points. *Science, 315*(5814), 972–976.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). Unsupervised learning. In *The elements of statistical learning* (pp. 485–585). New York: Springer.

Herscovics, N., & Linchevski, L. (1994). A cognitive gap between arithmetic and algebra. *Educational Studies in Mathematics*, *27*(1), 59–78.

Kieran, C. (1981). Concepts Associated with the Equality Symbol. *Educational Studies in Mathematics*, *12*(3), 317–326.

Kieran, C. (1985). The equation-solving errors of novice and intermidiate algebra students. In L. Streefland (Ed.), *Proceedings of the Annual Conference of the International Group for the Psychology of Mathematics Education* (pp. 141–146). Noordwijkerhout, The Netherlands.

Kieran, C. (1992). The Learning and Teaching of School Algebra. In D. Grouws (Ed.), *Handbook of Research on Mathematics Teaching and Learning* (pp. 390–419). New York: Macmillan Publishing Company.

Küchemann, D. (1981). Algebra. In K. Hart (Ed.), *Children's Understanding of Mathematics: 11-16* (pp. 102–119). London: Murray.

Linsell, C. (2007). Solving equations: Students' algebraic thinking. *Findings from the New Zealand Secondary Numeracy Project 2007*, 39–44.

Linsell, C. (2009). A Hierarchy of Strategies for Solving Linear Equations. In R. Hunter, B. Bicknell, & T. Burgess (Eds.), *Crossing divides: Proceedings of the 32nd annual conference of the Mathematics Education Research Group of Australasia* (Vol. 1, pp. 331–338). Palmerston North, NZ: MERGA.

Maaten, L. van der, & Hinton, G. (2008). Visualizing data using t-SNE. *Journal of Machine Learning Research*, *9*(Nov), 2579–2605.

Palm, T., Andersson, C., Boström, E., & Vingsle, C. (2017). A review of the impact of formative assessment on student achievement in mathematics. *Nordic Studies in Mathematics Education*, *22*(3), 25–50.

Pennington, J., Socher, R., & Manning, C. (2014). Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)* (pp. 1532–1543).

Rhine, S., Harrington, R., & Starr, C. (2018). *How Students Think When Doing Algebra.* Charlotte, NC: Information Age Publishing, Incorporated.

Vlassis, J. (2002). The balance model: Hindrance or support for the solving of linear equations with one unknown. *Educational Studies in Mathematics*, *49*(3), 341–359.

Paper B

Elkjær, M., & Jankvist, U. T. (2021). Designing tasks for a dynamic online environment: Applying research into students' difficulties with linear equations. *Mathematics*, *9*(5), 557. https://doi.org/10.3390/math9050557

mathematics

Article

Designing Tasks for a Dynamic Online Environment: Applying Research into Students' Difficulties with Linear Equations

Morten Elkjær * and Uffe Thomas Jankvist

Danish School of Education, Aarhus University, Campus Emdrup, DK-2400 Copenhagen NV, Denmark; utj@edu.au.dk
* Correspondence: me@edu.au.dk

Abstract: Despite almost half a century of research into students' difficulties with solving linear equations, these difficulties persist in everyday mathematics classes around the world. Furthermore, the difficulties reported decades ago are the same ones that persist today. With the immense number of dynamic online environments for mathematics teaching and learning that are emerging today, we are presented with a perhaps unique opportunity to do something about this. This study sets out to apply the research on lower secondary school students' difficulties with equation solving, in order to eventually inform students' personalised learning through a specific task design in a particular dynamic online environment (matematikfessor.dk). In doing so, task design theory is applied, particularly variation theory. The final design we present consists of eleven general equation types—ten types of arithmetical equations and one type of algebraic equation—and a broad range of variations of these, embedded in a potential learning-trajectory-tree structure. Besides establishing this tree structure, the main theoretical contribution of the study and the task design we present is the detailed treatment of the category of arithmetical equations, which also involves a new distinction between simplified and non-simplified arithmetical equations.

Keywords: students' mathematics-specific difficulties; linear equations; algebra; task design; dynamic online environments

Citation: Elkjær, M.; Jankvist, U.T. Designing Tasks for a Dynamic Online Environment: Applying Research into Students' Difficulties with Linear Equations. *Mathematics* **2021**, *9*, 557. https://doi.org/10.3390/math9050557

Academic Editor: Marta Molina

Received: 29 January 2021
Accepted: 27 February 2021
Published: 6 March 2021

1. Introduction

Dynamic online environments for teaching and learning mathematics are emerging all over the world. These online environments feature everything from video lectures to adaptive, gamified quizzes, and vary from simple training tasks to complex modelling tasks. A dynamic online environment is an online platform (or portal) where students may study mathematics in various ways. Initially, environments such as these consisted of just video lectures on various subjects or were gateways that provided access to a large number of training tasks to supplement the physical or electronic books used in the classroom. Recently, however, platforms have been developed where teachers and students interact on homework assignments, do differentiated group tasks, and so forth [1].

Dynamic online environments have the advantage of both offering access and having the capacity to process data provided in the form solutions to tasks. Such data may be, and is often advertised as, providing students with a highly streamlined experience in the environment, and an opportunity for a guided experience that leads to improved greater learning outcomes. Providing the user with an individualised learning path may be referred to as *personalised learning*. Personalised learning is an emerging concept that builds on the idea that no one, fixed learning path is suited to all learners [2]. However, according to Chen [2], one problem is that personalised e-learning systems tend to not consider whether a learner's ability and the difficulty level of the recommended courseware match each other. Chen [2] emphasises that learner ability should be considered when attempting to develop personalised learning. With regard to mathematics, learner ability may be interpreted in terms of the level of understanding of certain mathematical concepts related to a given situation.

To develop personalised learning paths in the field of linear equations, one must attempt to measure the student's ability or understanding of the concepts in this field. In this paper, we take a look at how dynamic online environments may benefit from the many years of mathematics education research through task design and collected data concerning students' work with linear equations, in order to promote personalised learning by determining learner ability when solving linear equations.

The task design principles presented in this paper yielded a collection of linear equations that target lower secondary school students. This collection of equations is intended to cover the intersection of the set of equations that may or ought to appear in the lower secondary school curriculum, and the set of equations highlighted in the literature as presenting difficulty to lower secondary school students. When addressing these design principles, we apply an adapted definition of the two types of difficulty that Jankvist and Niss associate with the concept of the linear equation [3]:

> 'The first kind of difficulty... is to do with goal-oriented transformation of equations (and, more fundamentally, algebraic expressions) into equivalent ones by way of permissible operations. [...] The second kind of difficulty, which appears to be of a more fundamental nature, is to do with what an equation actually is, and with what is meant by a solution to it.' [3] (p. 276)

By gaining a strong understanding of the prevalence of the concrete difficulties associated with seemingly common training tasks involving equations, one may find new possibilities related to both personalised learning and formative assessment. Feedback is a crucial part of task design in general [4]. The long term aim of the present design under discussion is to not only provide feedback on individual tasks, but also to offer the teacher additional refined feedback and auto-generated suggestions for how a student might improve their set of equation-solving strategies and knowledge related to the various mathematical concepts involving linear equations.

1.1. Matematikfessor.dk—A Dynamic Online Learning Environment

In Denmark, the dynamic online environment, *matematikfessor.dk* (Danish slang for 'mathematics professor') has existed for 10 years, and has approximately 500,000 student users who pay for access. To contextualise this, Denmark has a total of 700,000 students in primary school and lower secondary school, which means that over 70% of Danish school students have access to, and use the content on *matematikfessor.dk*. On an average day, 45,000 unique Danish students use the variety of tasks offered by the site, and collectively provide solutions to 1,500,000 tasks. This means that on an average day, almost 10% of the user base accesses this online environment. Every day, 45,000 of the 500,000 student users complete 33 tasks, on average. During the average Danish school year, approximately 300,000,000 tasks are completed on *matematikfessor.dk*. To illustrate how widespread the use of this environment is, when Denmark was locked down on 11 March 2020, owing to COVID-19, the number of tasks completed daily rose to approximately 6,000,000. The number of unique daily visitors on a standard weekday was approximately 45,000–50,000. The number of unique daily logins rose to 130,000.

The dynamic aspect of the platform consists in the platform's ability to learn certain things about the students. When a student provides a solution to a task, the system (an algorithm) may present the student the next task based on this solution (as well as previous solutions). As an example, the equations presented in this study could be assigned to a student in a dynamic manner. This is done by the platform (system) learning what types of equations and which variations the student is able to solve. On the one hand, the system can present the student with tasks that are challenging in relation to level and performance. On the other hand, the system can gather information about the student's ability as an equation solver, and thereby potentially inform the student's teacher in relation to future learning trajectories. This dynamism is what leads to the enabling of personalised learning.

Dynamic online environments are subject to several obvious characteristics. The platform's content may dynamically change; new tasks may be added to the environment and old tasks may be removed or updated. This provides the editors of online environments with the added flexibility of dynamically adapting it to the needs of its user base. These parameters give online environments a clear advantage over printed teaching material materials such as textbooks. Another characteristic of online learning environments is that teachers may assign tasks from the online environment to the students individually, including deadlines, monitoring the solutions and providing feedback. Many online environments also provide the students with auto-generated feedback and suggestions for how they could have completed a given task, for example, if a wrong solution is submitted. This feedback may vary, and may provide the student with the correct solution to a detailed explanation of how the task should (could) be correctly completed.

Although some of the disadvantages of using dynamic online environments may be obvious, others are more hidden. A teacher may have certain beliefs regarding how a subject and its concepts should be explained or taught. When presenting video lectures in an online environment, one may fear that students are merely presented with algorithms that solve a textbook problem, and not with a general strategy for solving a whole class of problems. Other—perhaps less obvious—disadvantages are the consequences of particular task designs. As the online environment is expected to provide instant feedback, the user is allowed only certain input types. Some problems would require a student to justify an assumption or in some way provide a string of text, or even spoken words. Most of the tasks presented in *matematikfessor.dk* are what many would refer to as training tasks. Therefore, in most cases the two types of input that students may provide are either responses to multiple-choice questions, or numbers. In task design, this becomes a significant disadvantage, particularly when working with algebraic expressions. It means that it is not possible to prompt students for an algebraic expression without presenting the student with multiple choices. One might argue that this is acceptable in many cases. Nevertheless, it forces upon the task designer an additional objective, namely, to create *distractors*. Preferably, such distractors should be chosen or designed based on scientific or experiential results that qualify the answer as a good distractor. One could argue that for a student, a poorly designed wrong solution would make a task easier or 'bad'—and in most cases, more problematic—in terms of conducting research.

1.2. Research Question

The preceding discussion led us to the following research question:

> *How may research on lower secondary school students' difficulties with linear equations inform task design in a dynamic online environment with the possibility to promote/support students' personalised learning?*

2. Methodology Part 1: Finding Key Publications

With regard to conducting a review of the literature concerning the difficulties lower secondary school students encounter when working with linear equations, the reader should be aware of the sheer volume of the body of mathematics education literature concerning students' difficulties with algebra. Although this review mainly seeks to uncover the difficulties that arise when students work with equations and attempt to solve them, many of these difficulties have a natural relationship to the subject of algebra. Hence, it became clear that a strategy focused solely on a database search would be neither sufficient nor possible to accomplish. Instead, the literature review relies on parts of the hermeneutic approach [5]. This means that the strategy was *not* to apply a systematic approach that covers all the literature on this subject, but to use initial database search iterations, and handbooks and existing literature reviews, to pave the way for a suitable citation-tracking—snowballing—approach to identify key publications that address the difficulties encountered when working with equations [5]. Several major literature reviews on children's difficulties with algebra have been conducted over the years. Snowballing based on database searches made it possible to identify a number of extensive literature reviews on difficulties with algebra. Similarly, it was possible to identify key publications that described students' difficulties with linear equations in chapters in the leading handbooks on mathematics education. Finally, the advice and knowledge of experts in the field of mathematics education has been valuable.

The database searches were carried out using the MathEduc (note that MathEduc is no longer operational: https://www.zentralblatt-math.org/matheduc/, autumn 2018) and ERIC (https://eric.ed.gov/) databases, with various search strings containing keywords such as 'equation', 'arithmetic', 'algebra', 'misconception', 'difficulties' and 'error', which, after culling out duplicates, yielded 232 items. After studying their titles and abstracts, 68 items were found to be relevant to our study. These studies are listed in Appendix A. Citation tracking was applied to the remaining references and to a search for books and comprehensive works on difficulties with algebra. For example, Rhine, Harrington and Starr's book [6] includes over 900 works of reference covering the last five decades of research on learning algebra. Their work and others'—such as Booth et al. [7], Jupri, Drijvers and van den Heuvel-Panhuizen [8], Kieran [9,10] and Booth [11]—paved the way for practical citation tracking (snowballing) [5].

In the following, we present the findings of our study of the literature. As mentioned, these findings are restricted to information that describes difficulty in understanding and solving equations (particularly among lower secondary school students), which may inform task design related to linear equations in dynamic online environments. As part of this, we also present and analyse existing didactic categorisations of linear equations and key difficulties that students encounter when working with linear equations.

3. Key References from Five Decades of Research

We have organised our findings (or students' difficulties) into the following categories and subcategories: (1) existing categorisations of linear equations; (2) difficulties related to the concept of 'number'; (3) difficulties related to interpretations and the role of the equals sign; (4) strategic and transformational difficulties; (5a) operations and conventions; and (5b) letters and expressions.

3.1. Existing Didactic Categorisations of Linear Equations

Filloy and Rojano [12] argue for the presence of a 'didactic cut', which appears when the unknown in an equation is present on both sides of the equals sign, in which case the students must work with the unknown. Due to this didactic cut, Filloy and Rojano [12] separate linear equations into the following two categories:

- Arithmetical equations: $ax + b = c$
- Non-arithmetical equations: $ax + b = cx + d$

Although quite useful as a preliminary rough approximation, Filloy and Rojano's distinction is a little too simple (e.g., [13]), when it comes to all the different types of linear equations that one can possibly imagine. The didactic cut argument is that students with no prior instruction in solving equations can solve those of the general $ax + b = c$ type, but are unable to solve equations of the $ax + b = cx + d$ type. Still, Filloy and Rojano's [12] distinction deserves attention.

Finding additional categorisations of linear equations in the literature proved to be a challenge. However, Vlassis [14] does present an extension of Filloy and Rojano's distinction. As others have, Vlassis [14] challenges this distinction. She argues that both arithmetical- and non-arithmetical equations should be further divided into two categories each, where a, b, c and d are positive integers:

- Arithmetical equations
    - Concrete arithmetical equations: $ax + b = c$
    - Abstract arithmetical equations: $ax + b - cx = d, -x = c$
- Non-arithmetical equations

- ○ Pre-algebraic equations: $ax + b = cx + d$
- ○ Algebraic equations: $-ax = cx - d$,

The justification for this separation is based on the argument that not all arithmetical equations may be solved with an arithmetical approach to the expression. Vlassis argues that equations such as $-x = 7$, or even $6x + 5 - 8x = 27$, fall into the arithmetical equation category, but cannot be solved entirely with arithmetic alone. Hence, the need for the abstract arithmetical equation category. In the case of the non-arithmetical equations, Vlassis argues that the presence of negative numbers (in coefficients and terms) causes the main detachment from a use of a model, e.g., the balance model, to solve or interpret an equation [14].

3.2. Difficulties Related to the Concept of the Number

Working with negative numbers is a major theme when addressing students' difficulties in working with linear equations [14–18]. In arithmetic, the negative sign indicates only the operation of subtraction. However, in algebra the negative sign has new predicative purposes, namely unary, binary and symmetrical [18]. The binary role is one that lower secondary school students may know from carrying out arithmetic calculations, whereas in algebra, they are presented with additive inverses (symmetrical role) and the minus sign as a predicate to a number (unary role). It is argued that beginning algebra students find the negative sign's new role counterintuitive. Thus, the difficulties the students encounter include both negative numbers and the separation of the sign from the number or letter [13,17]. Gallardo [16] presents four levels of interpretation of negative numbers:

1. *Subtrahend*, where the notion of number is subordinated to the magnitude (for example, in $a - b$, a is always greater than b, where a and b are natural numbers).
2. *Relative or directed number*, where the idea of opposite quantities in relation to a quality arises in the discrete domain and the idea of symmetry appears in the continuous domain.
3. *Isolated number*, that of the result of an operation or as the solution to a problem or equation.
4. *Formal negative number*, a mathematical notion of negative number, within an enlarged concept of number embracing both positive and negative numbers (today's integers). (p. 179)

Kieran [19] argues that two overarching difficulties emerge when solving equations with negative numbers, namely:

- Inversing a subtraction with a subtraction or failure to do so when necessary, e.g., solving $16x - 215 = 265$ by subtracting 215 from 265 or solving $37 - b = 18$ by adding 37 and 18. (p. 143)
- Leaving the unknown with a negative sign in front of it, e.g., $-x = -17$. (p. 144)

This leads to the observation that negative numbers, the predicative and operation minus, are of paramount importance when we consider students' difficulties with various types of linear equations. Negative numbers are not only difficult to work with, as is mentioned later in the section on transformational activity, but they are also difficult to work with when they are the solution to a problem [6,16,18,19]. Christou et al. [20] emphasise that the unknown value may stand for either a positive or a negative number, independently of the sign that is attached to it.

Negative numbers are not the only ones that present difficulties for students. Fractions, as coefficients and as solutions to equations, also do. Christou and his colleagues [20] point out that for the students, fractions may be 'unexpected' solutions to linear equations. This may partly be understood with reference to the didactic contract [21], in particular when none of the numbers in a given equation are rational (fractions, decimal numbers, etc.) (e.g., [3,22]); it may also be the result of the common misconception that multiplication always makes a number 'bigger' [20], or of common misinterpretations of letters that we will touch on later.

Research suggests that students have difficulties when working with, or encountering, the number 0. In a literature review, Rhine et al. [6] present a table of various understandings of 0 in mathematics:

- *Nil zero*: Zero has no value and students act as if it was not there.
- *Place value zero*: Zero is used as a placeholder in a large number when there is none of that place value (over half of students up to the eighth grade could not write a number such as 'two hundred thousand forty three').
- *Implicit zero*: The zero does not appear in writing, but is used in solving a task. A student might solve a problem by thinking about obtaining a zero in the process. For example, $5 - 17 = 5 - 5 - 12 = 0 - 12 = -12$ might be the thought process while $5 - 17 = -12$ is the only thing written.
- *Total zero*: The combination of number opposites. For example, $34 + (-34) = 0$.
- *Arithmetic zero*: The result of an arithmetic operation.
- *Algebraic zero*: The result of an algebraic operation or the solution of an equation. (p. 145)

The foregoing understandings are also related to the concept of the equation, especially when it comes to equation-solving or solutions to equations that are 0. Rhine et al. [6] also emphasise that students often are confused by the appearance of the number 0 in equations, and mistakenly believe that the solution to the equation is 0 (e.g., $0x = 5$).

3.3. Difficulties Related to the Equals Sign

In the cohort of publications we selected for our study, several research studies report findings that support the observation that students' understanding of the equals sign has a tremendous impact on their success with algebra. Studies also report that this is true of success with equation-solving, mainly in relation to the ability to carry out the same operation on either side of an equation [13,23–27]. One of the central difficulties that students encounter in the cognitive transition from an arithmetic approach—making calculations—to an algebraic approach—simplifying expressions and working with equations—is that they continue to view the equals sign as a '"do something" signal' [25,28], or they maintain an urge to 'calculate', out of habit [29,30].

Behr, Erlwanger and Nichols [31], and Kieran [28] reported on students who viewed expressions such as $\square = 2 + 3$ as being 'backwards'. Such findings suggest that some students look for the opportunity to 'calculate' or to get 'the solution, and that decomposing a number is an acquired way of thought that they bring from arithmetic. This links to another phenomenon, namely, accepting a lack of closure, which effectively means not accepting an expression as 'the solution or an expression on the right side of the equals sign [32–34]. For example, Linsell [35] presents data that shows that the equation, $26 = 10 + 4n$, is much more challenging for students than the seemingly similar equation, $4n + 9 = 37$.

Beginning algebra students tend to believe that only a single entity should be present on the right side of the equals sign, which, in Falkner, Levi and Carpenter's [23] famous type of equation, $8 + 4 = \square + 5$, means that students often believe that 12 must go in the box, as they simply calculate from left to right. For this same equation, some students believe that a 5 should be added to the 12, so 17 goes in the box, as they desperately attempt to keep the 'answer' free of operators. In their study, only 2% of 5th and 6th grade students gave the correct answer, that is, 7 [36]. This example demonstrates that students may be willing to ignore the meaning of the operator's position in their efforts to give the proposition meaning [37]. This also demonstrates some students' willingness to disregard the last term, to keep the equation in an 'operation equals answer' form [28].

In another example, students may believe that in the equation, $4 + 5 = 3 + 6$, the term $3 + 6$ should be combined into the 'solution, owing to its position to the right of the equals sign. This is also the case with a similar problem involving an algebraic expression that includes letters as 'the answer' [11,38]. The way such a view is linked to the concept of the equation may prove rather problematic. If students' perception of the equals sign is associated with the button on a calculator that indicates that a calculation is made, or must take place, they are clearly still in an arithmetic mindset, and will experience difficulties when working with certain types of arithmetical and non-arithmetical equations.

Sfard and Linchevski [39] address students' incorrect use of strings of equations. For instance, when handling a word problem such as, 'How many marbles do you have after you win 4 marbles 3 times and 2 marbles 5 times?', the child will often write: '$3 \cdot 4 = 12 + 5 \cdot 2 = 12 + 10 = 22$' (p. 209). They emphasise that such a use of the equals sign clearly demonstrates an arithmetic or process-oriented mindset. These arguments suggest that the position of the equals sign and the structure of the right side of the equation play major roles in a didactic categorisation of various linear equations, according to students' difficulties with such.

Matthews et al. [27] suggest the following levels of understanding of the equals sign (see Table 1).

Table 1. 'Construct map for knowledge of the equals sign as an indicator of mathematical equality.' (adapted from [27] (p. 224)).

Level 4: Comparative Relational	Successfully solve and evaluate equations by comparing the expressions on the two sides of the equal sign, including using compensatory strategies and recognising transformations maintain equality. Consistently generate a relational interpretation of the equal sign.	Equations that can be most efficiently solved by applying simplifying transformations: For example, without adding 67 + 86, can you tell if the number sentence '67 + 86 = 68 + 85' is true or false?
Level 3: Basic Relational	Successfully solve, evaluate, and encode equation structures with operations on both sides of the equal sign. Recognise relational definition of the equal sign as correct.	Operations on both sides: $a + b = c + d$ $a + b - c = d + e$
Level 2: Flexible Operational	Successfully solve, evaluate, and encode atypical equation structures that remain compatible with an operational view of the equal sign.	Operations on right: $c = a + b$ No operations: $a = a$
Level 1: Rigid Operational	Only successful with equations with an operations-equals answer structure, including solving, evaluating, and encoding equations with this structure. Define the equal sign operationally.	Operations on left: $a + b = c$ (including when blank is before the equal sign)

These levels suggest that a student should at least have a 'flexible operations' conception of the equals sign, to be able to solve, manipulate or possibly even understand (decode [40]) algebraic equations [14].

3.4. Difficulties Related to Strategies and Transformations

Jankvist and Niss [3] mention two major concerns that are at stake with regard to the difficulty related to goal-oriented transformations of equations into equivalent expressions. First, there is the difficulty of choosing a sequence of efficient operations that leads to transformations of the equation in question, and gives the student an opportunity to find an equation from which the solution may be inferred. Second, there is the difficulty of making valid transformations of the elements in the equation. Several studies show that for students, the difficulty of an equation increases with the number of steps required to solve it (with the formal approach) (e.g., [41–43]). For example, Rhine et al. state [6]:

> One- and two-step equations represent classes of algebraic relations that are often the easiest for students to solve. These equations often take the form of $ax + b = c$ where a, b and c are constants. Students may visualise what has been 'done to' and 'undoing what has been done' as an informal way of approaching the task that is often successful.
>
> (p. 100)

According to Araya et al. [44], when solving arithmetical equations, students must solve equations with the unknown on the right-hand side of the equals sign instead of the left. This is consistent with Behr et al.'s findings [31] that students may view certain equations as 'backwards', and with Linsell's [35] mentioning that $4n + 9 = 37$ is less challenging for students than 26 = 10 + 4n. Furthermore, Araya et al. [44] mention that changing the letter used to represent the unknown, such as 'x', for a less commonly used letter, such as 's', makes an equation more difficult for some students, since they do not have a strategy for 's'. Additionally, constructing equations with an infinite number of solutions and equations with no solutions are further possibilities for distinction. In a study of 621 lower secondary school students, Linsell [35] measures their ability in terms of the equation-solving strategies they apply, and in this way establishes a hierarchy of equation-solving strategies. Linsell mentions that the strategies applied varied from equation to equation, indicating that some strategies are only applied in certain situations involving certain properties of the equations. For example, Linsell [35] mentions that the equation, $5n - 2 = 3n + 6$, was solved using only a guess and check approach, or formal transformations.

Most of the strategies listed in Table 2 are more or less self-explanatory, except perhaps for the 'cover-up' strategy. This strategy or technique is based on the idea of working backwards, and substitution. The student covers the term (or expression, in some cases) containing the unknown, and solves the equation using the 'cover' as a new unknown. Next, the student solves for the initial unknown by setting the solution found for the cover 'equal to' the cover. For example, one could cover the term $6x$ in the equation, $6x + 5 = 23$, to get $\square + 5 = 23$, realising that $6x = 18$, before solving for x in this new equation.

Table 2. Classification of strategies used for coding [35] (p. 333), drawing on [10].

1.	Unable to answer question
2.	Known basic facts
3.	Counting techniques
4.	Inverse operation
5.	Guess and check
6.	Cover up
7.	Working backwards then guess and check
8.	Working backwards then known fact
9.	Working backwards
10.	Transformations/equation as object

Strategy no. 10 (Table 2) is also referred to as the 'formal strategy'. Kieran elaborates that students bring the two techniques of 'using a known basic fact' and 'the counting technique' with them from arithmetic, to solve 'missing addend sentences'. Even though we cannot necessarily list these strategies in a hierarchy of sophistication for all linear equations, Linsell's findings suggest that the various types of linear equation do have some hierarchy [35]. The meaning a student assigns to the equals sign or to the equation as a concept is a key to choosing a sequence of transformations that will lead to the solution to an equation. It should be noted here that an algorithmic approach to equation-solving could lead to the correct solution without the student having a proper understanding of the concepts and theorems involved. The second aspect of the difficulty presented by transformational activity [3] involves the valid manipulations of the elements in an equation, both the actions that should maintain the equality of the expressions and the manipulations that take place in the expressions on either side of the equals sign. This aspect of the transformational difficulties may bear a closer relationship to difficulties with arithmetic and algebra in general. Kieran [19] describes a range of common errors children make in their attempts to solve a linear equation. Some errors are made by beginners, some by intermediates, and some errors are made by both groups. Errors common to both groups were:

- Giving up when attempting to solve using the substitution procedure.
- Inversing subtraction with subtraction and addition with addition

- Computing a coefficient with a non-coefficient
- Forgetting that concatenation means multiplication (p. 143)

When students learn to solve equations, the 'change side—change sign' strategy may be a powerful tool for quickly progressing and solving equations rapidly. However, this strategy may lead the students to commit errors 2 and 3 [6]. Among beginning algebra students, observed difficulties included [19]:

- Not using the order of operations convention.
- Not knowing how to start solving a given equation-type.
- Inversing a multi-operation equation before collecting the multiplicative terms.
- Not using the convention that two occurrences of the same unknown are the same number.
- Giving precedence to an addition when it is preceded by a subtraction.
- Inversing a two-operation equation only once and then using the result of that operation as the solution. (p. 144)

Observed errors committed by intermediate algebra students included [19]:

- Leaving the unknown with a negative sign in front of it, e.g., $-x = -17$.
- Changing an addition to a subtraction when transposing, but then commuting the subtraction, e.g., $30 = x + 7 \rightarrow 7 - 30 = x$.
- Transposing only the literal part of the term and leaving the coefficient behind, e.g., solving $7 \times c = c + 8$ by writing $7 - 8 = c \div c$.
- Dividing larger by smaller rather than respecting the order for inversing, e.g., $11x = 9 \rightarrow x = 11/9$.
- Computational error involving positive and negative numbers.
- Inversing a one-operation addition equation twice by inversing the addition and then dividing the unknown by the result of the subtraction, e.g., solving e.g., solving $n + 6 = 18$ by subtracting 6 from 18 and then attempting to divide n by 12. (p. 144)

Prediger [34] argues that "If, in the calculation aspect, the variables are considered to be meaningless symbols, terms are also meaningless expressions and equivalent terms are those which can be transformed into each other according to the transformation rules." [34] (p. 7).

3.5. Difficulties with Letters in Expressions of Linear Equations

When examining the literature related to students' difficulties with interpreting and working with letters in mathematical expressions, many researchers report findings related to algebra in general [6,9,10]. The majority of these studies are not specifically about equations or equation-solving. Usiskin [45] suggests that the concept of algebraic variables may be seen in four different ways: generalising arithmetic; solving equations, including making sense of solutions; exercising algebraic rules while studying relationships among quantities; and building algebraic structures and learning from these. Usiskin [45] raises the point that:

> Under the conception of algebra as a generalizer of patterns, we do not have unknowns. We generalize known relationships among numbers, and so we do not have even the feeling of unknowns.
>
> (p. 12)

Usiskin elaborates that working with variables as unknowns is different from working with variables when generalising arithmetic and when working with variables in formulas.

When working towards a solution to an equation, students assign different meanings to the letters in it. During the CSMS project [46], Küchemann [38,47] developed six categories of views that students apply to letters in their attempts to make sense of algebraic expressions. These categories are mentioned extensively in the literature on children's thinking about algebra, yet, they should not be interpreted as hierarchical levels (e.g., [6]). In Küchemann's first three categories, we find student difficulties that relate to working with unknowns in equations. We address these categories one by one.

In the 'letter evaluated' category, students will avoid operating on the letter as an unknown by assigning a value to the letter. An example would be assigning an 'alphabetical value' to the letter. This value corresponds to the numerical value based on the letters' positions in the alphabet. Similarly, students may assign the letter the value 1, because there is one 'thing' present. Lastly, students may assign the letter a contextual value drawn from the given situation [38]. In the best-case scenario, adopting this view of the unknown letter when solving a linear equation may be successful, as long as no operations on the unknown are required.

In the 'letter not used' category, students ignore the letter, or at best, acknowledge its existence without ascribing a meaning to it. This category is exemplified by solving this equation: if $a + b = 43$, $a + b + 2 =$____. In this example, the students do not need to assign values to the letters a and b to reach a solution. However, consider the following example: if $e + f = 8$, $e + f + g =$____. Although the same view may be applied here, acceptance of the lack of closure is required. In the so-called CSMS study, this led students to evaluate the letter g in order to give answers such as $4 + 4 + 4 = 12$ or 15, because g is the 7th letter of the alphabet. Such alphabetical evaluation has also been reported by MacGregor and Stacey [48]. Both Kieran [19], Küchemann [38], and Booth [11] report on students who compute coefficients with non-coefficients (also known as conjoining). In an equation, this could result in $3x + 7$ being substituted with $10x$.

In the third, 'letter used as an object', category, students regard the letter as shorthand for an object, or as an object in its own right. This view of letters used in algebra may prove beneficial in some cases, but disastrous in others. If this view is applied in order to collect like terms, treating the letters as objects works well. For example, when simplifying $2x$—3 + x to $3x - 3$, the x's may be thought of as bananas, chocolate bars or just x's. However, some students tend to treat the letter as an abbreviated word, such as 'height' or 'blue', instead of merely being a placeholder for a number. This view of the letters corresponds with the idea of the equation fitting a model, for example, the balance model [14].

After presenting the three, above-noted original categories, Küchemann [38] presents three additional ones, namely, 'letter as a specific unknown', 'letter generalised number' and 'letter used as variable'. In each of these three additional categories, students may operate on the unknown number in such a way that it facilitates the solution of an equation. Hodgen et al. [49] argue that students' understandings of letters in mathematics have not changed over the years, that is, not since Küchemann first described these.

4. Methodology Part 2: Task Design

In this section, we present the task design principles and the reasons behind them. When designing tasks for any environment, one should be aware of the main purpose of developing these tasks, since they may serve several purposes. Burkhardt and Swan [50] present four general categories of task design: specifying a curriculum; high-stakes assessment; classroom assessment; and teaching and learning. Burkhardt and Swan [50] mention that high-stakes assessment tasks may play three roles, namely: measuring performance; exemplifying performance goals; and driving classroom learning activities. Notably, the task design in this project covers a range—or a sequence—of tasks that are not meant to be evaluated or assessed individually. To some extent, our sequence of tasks resembles high-stakes assessment, in that it may be seen as measuring performance. However, as mentioned in the introduction, the overarching purpose of this project is to offer a formative assessment of what may initially resemble regular, repetitive training tasks.

Applying variation theory [51] to a concept such as the linear equation makes it possible to carefully—and meaningfully—design a sequence of tasks that may be seen as a whole [52]. By varying the construction of each task in a sequence to address the difficulties presented in the preceding section, it becomes possible to develop meaningful tasks that may be evaluated as a set. In variation theory you work with the *intended object of learning*, which is what the teacher or the task intends to demonstrate to the student; and the *lived object of learning*, which is what actually becomes the student's focus, when he/she executes a task. The assumption is that if the tasks are carefully constructed, the intended object will become obvious under an assessment. The goal is to make the intended object and the lived object coincide.

When specifying a sequence of tasks nearly similar to each other for online use, Bokhove [53] suggests *the element of crisis* as an important factor. Such a crisis occurs when the student completing a range of tasks—or parts of the task range—encounters a task that is impossible, or nearly impossible, to solve. This element of crisis resembles a cognitive conflict (e.g., [54]) or an inadequate conceptual field [55]. Bokhove and Drijvers [4] use the element of crisis in their variations, when designing a sequence of nearly similar tasks to show that students attempting to solve a crisis-provoking task similarly to pre-crisis tasks may result in incompleteness, because the earlier strategy is inadequate.

Burkhardt and Swan [50] suggest that the question of task difficulty must not be ignored, and is an important part of task classification and design. They propose four considerations related to task difficulty:

- Complexity—the number of variables, the variety and amount of data, and the number of modes in which information is presented, are some of the aspects of task complexity that affect the difficulty it presents.
- Unfamiliarity—non-routine tasks (those which are not just like the tasks one has practised solving) are more difficult than routine exercises.
- Technical demand—tasks that require more sophisticated mathematics for their solution are more difficult than those that can be solved with more elementary mathematics.
- Student autonomy—guidance from an expert (usually the teacher), or from the task itself (e.g., by structuring or 'scaffolding' it into successive parts) makes a task easier than if it is presented without such guidance. [50] (p. 433)

Watson and Mason [52] propose the *controlled variation* of tasks, which offers the following elements:

- Analysis of concepts in the conventional canon that one hopes learners will encounter.
- Identification of regularities in conventional examples of that concept (and its related techniques, images, language, contexts) that might help learners (re)construct generalities associated with the concept. Even an algorithm can be seen as a generality.
- Identification of variation(s) that would exemplify these generalities; decide which dimensions to vary and how to vary them;
- Construct exercises that offer micro-modelling opportunities, by presenting controlled variation, so that learners might observe regularities and differences, develop expectations, make comparisons, have surprises, test, adapt and confirm their conjectures within the exercise;
- Provide sequences of micro-modelling opportunities, based on sequences of hypothetical responses to variation, that nurture shifts between focusing on changes, relationships, properties, and relationships between properties. [52] (pp. 26–27)

An Example Task

As an example, let us take a look at the general arithmetical equation, $ax + b = c$ [12], where a, b and c are all positive integers. In many cases, c would be a larger number than b, to keep the value of x (the solution) positive. However, bearing in mind that students often have difficulty with negative numbers, a variation might look like this:

$$4x + 8 = 4$$

The values of a, b and c are chosen so that a is a factor in $c + b$, $c - b$ and $b - c$. The solution to an equation such as the example above takes the form of $x = (c - b)/a = (4 - 8)/4 = -1$. If we recall that lower secondary school students may not be comfortable with negative numbers, and therefore unwilling to proceed by rearranging the numbers when subtracting the $4 - 8$, they may end up with 4 instead of -4, and thus reach the incorrect solution of $x = 1$. To address another common difficulty, inverting addition by addition, some students may add b and c, and find $x = 3$. Finding the wrong answer to an equation may be due to unfamiliarity with this sort of equation, so a negative number is found for the solution [50]. This may also lead to reversing the order of the subtraction, resulting in $b - c$, to make the task make sense, and the same student might have subtracted b from c if $c - b$ yielded a positive number. Some may describe this as a result of breaking the didactic contract [21] between the teacher (issuing the task) and the student. The goal of the design is to increase complexity, unfamiliarity and technical demands for the range of tasks, while using variation theory to present subtypes that vary to address the range of difficulties reported in the literature.

When working towards the foregoing general design goals, some minor detours will be made, to address specific difficulties (explained in the next section). The intentional crisis, combined with the knowledge of major difficulties that students encounter when solving linear equations and their knowledge of useful strategies, becomes the strategy used to design a broad range of increasingly complex, unfamiliar, technically demanding linear equations, in order to assess online students' abilities to solve linear equations.

5. Establishing Categories of Linear Equations

In this section, we explain our task design of the various types of linear equations, based on the results of research into students' difficulties with linear equations, and the task-design principles guided by variation theory. More precisely, the task design takes its point of departure in the categorisation established by Filloy and Rojano [12], and Vlassis's later expansion [14], bearing in mind that Vlassis' categorisation [14] is based on the problems and difficulties encountered when linear equations are made abstract by a negative sign/number. When separating linear equations into arithmetical and non-arithmetical (algebraic) equations, the reader should also bear in mind that the term *simplified arithmetical equation* indicates arithmetical equations with a basic arithmetic form, that is, $ax + b = c$ [12]. Similarly, the term *simplified algebraic equation* indicates algebraic equations with the form, $ax + b = cx + d$.

Our design included a set of boundaries. First, we wanted to make sure that we had a clear framework for separating the equations, or types of equations, which were included in the design. When we say 'separating', we mean a separation in terms of the difficulty and complexity of solving an equation, supported by our findings in the literature [50]. Second, to the extent possible, we set out to design a comprehensive arsenal of tasks for beginning or early intermediate equation solvers (e.g., Danish lower secondary school students in their 7th year). Bear in mind that, on the one hand, these tasks should present a learning path that also serves as a tool for detecting students with difficulties, or in the best-case scenario, a tool for showcasing students' concrete (categories of) difficulties with linear equations. Each type of equation in the design includes several variations of equations derived from the general type. In the next section, we further specify the design requirements.

5.1. Overarching Design Requirements and Goals

We remind the reader of the overall variations due to students' difficulties identified through the literature, task design theory, and in particular, the restrictions that exist when designing tasks for dynamic online environments. The intent of our design is to vary across:

- Negative numbers and the minus sign
  - as solutions
  - as terms
  - as operations
- Rational numbers
  - as solutions
  - as present numbers
- Interpretation of the equals sign

    - ○ Situations that invoke an element of crises from an arithmetic point of view
- Strategic, conventional and transformational questions
    - ○ Increasingly complex and strategically demanding
    - ○ Conventions concerning brackets
    - ○ Conventions concerning missing multiplication

As mentioned previously, student responses to tasks in the dynamic online environment that underpins our design are restricted to multiple-choice and input fields. For our design, the 'input field' was chosen as the only option. This was done to avoid influencing the students' answers, as may occur when presenting multiple choices. This, and the concept of the element of crisis [4], establishes an outset for applying variation theory [51], when designing and developing the tasks for the online environment.

5.2. *Arithmetical Equations*

Even though arithmetical equations [12] may appear simple, the smallest variations may be deal-breakers for some students. Small variations in, and considerations of task design, lead to tasks better suited to addressing the difficulties mentioned in the literature. This should not be seen as an attempt to trick the students into giving incorrect solutions, but as exposing a tendency to apply inadequate schemes when completing the task. Every type of simplified arithmetical equation (see Figure 1) is categorised into subtypes (or variations), using variation theory related to students' difficulties mentioned in the literature.

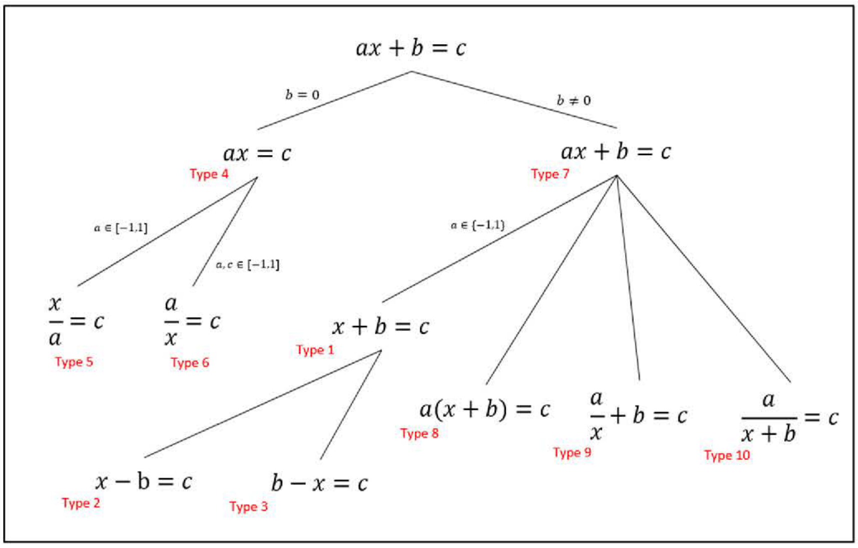


Figure 1. Simplified arithmetical equations.

The reader should note that in several of the variations, constant values (a, b and c) are simply switched, or in some cases one value is simply inverted, to establish a connection between the tasks, thereby adding another layer of possible comparison.

We now address and discuss each of the ten different types of arithmetical equations.

5.2.1. Types 1 and 2: $x + b = c$ and $x - b = c$

In many cases, these types of equations are what many teachers and students consider the simplest setup possible, because addition is arguably the simplest and most familiar operation to many students. All these equations are designed around the $x + b = c$ template, with various b and c values. Technically speaking, type 2 is simply a variation on type 1. For the sake of clarity, these are separated into two types. Although this form does not necessarily yield the simplest equations, many would consider several forms of the equations in this category easy or beginner level.

The two types of variations of these equations address the negative (for the values b, c and x) and the order of operations in the expression. To introduce a rational solution to these equations, one would have to let either b and/or c be a rational number. This was not deemed a necessary variation at this stage. The first subtype consists of simple addition and simple subtraction expressions, with small natural numbers for the b and c values. These are meant to be simple and easy for most upper secondary school students to solve. A solution typically requires a 'guess and check', a counting technique or the 'knowing' strategy, or a formal transformation, subtracting b from c, and isolating x. These equations add to the overarching task design by serving as a reference point. Table 3 below presents examples from this set of tasks. Note that the b and c values are exchanged whenever possible, to facilitate further a potential comparison of students' task performances.

Table 3. Examples of type 1 and type 2 equations and variations.

Original	Variation 1	Variation 2	Variation 3	Variation 4	Variation 5
$4+x=7$	$7=x+4$	$7+x=4$	$4+x=-7$	$7+x=-4$	$x+4=7$
$x-3=5$	$5=x-3$	$x-5=3$	$x-3=-5$	$x-5=-3$	$-5+x=3$

Variation 1 consists of simply flipping the equation so the addition is on the right hand side of the equals sign. This variation was chosen because of students' perceptions of the equals sign described in the literature. Variation 2 has a negative x value (solution), because the b and c values are exchanged. Variation 3 also has a negative x value, since the value of c is an additive inverse. The reason for including variation 3 is to try to determine whether the presence of a negative c value has any influence on the task performance, in comparison to variation 2. In variation 4, the b and c values are again exchanged, but the minus sign remains on the right side. This has no major influence on the tasks, where x is added to b. However, it changes the value of x from negative to positive in the tasks where b is subtracted from x. In variation 5, the order of operations on the left side is exchanged, and the value of x remains positive.

5.2.2. Type 3: $b-x=c$

This type of equation differs from types 1 and 2 in technically being a two-step equation, if solved formally or through formal transformations. Still, in many cases these equations are presumably solved by using 'guess-and-check', 'knowing', or counting strategies. For many students, an example equation would read as '6 less some amount makes 3'. If such equations are solved formally, the students are faced with an equation that looks like $-x=p$, where $p=c-b$. This type of equation is mentioned by Vlassis [14] as being one of the certain indicators of and reasons for finding that difficulty is not determined merely by whether an equation is arithmetical or non-arithmetical, but also by the level of abstraction it involves.

The two variations that appear in these equations are the use of the negative (for the values b, c and x), and the changes in the order of terms in the expression. Examples are given in Table 4.

Table 4. Examples of type 3 equations and variations.

Original	Variation 1	Variation 2	Variation 3	Variation 4	Variation 5
$6-x=3$	$3-x=6$	$-6-x=3$	$6-x=-3$	$-6-x=-3$	$-x+3=6$

The first variation exchanges the b and c values. This results in negative x values, when compared to the original, which has a positive x value. In variations 2, 3 and 4, b and c values are exchanged for their negative counterparts. In variation 5, the equations actually become a special case of type 7 (see Figure 1) equations. This variation is intentionally presented here.

5.2.3. Type 4: $ax = c$

This type of equation is the first, and technically simplest, version of an arithmetical equation that involves multiplication, with $b = 0$. Presumably, as with the previous types, many students will be able to apply a 'guess-and-check' or 'knowing' strategy, or a 'cover up' strategy when working with these equations. Using small positive integers ensures that the solution is also a natural number. In the standard or original variation, the equation may remind students of multiplication tables. Examples are provided in Table 5.

Table 5. Examples of type 4 equations and variations.

Original	Variation 1	Variation 2	Variation 3	Variation 4	Variation 5
$3x = 15$ $8x = 8$	$3x = -15$	$-3x = 15$	$-3x = -15$	$15 = 3x$	$28x = 14$ $2x = 9$

Variation 1 introduces a negative c value, which results in a negative x value (solution). The same idea is present in variation 2, where the a value is negative. In variation 3, the x value is again positive, but both a and c values are negative. Variation 4 introduces what was also introduced in type 1 (see Figure 1), where the expressions flip sides. In variation 5, rational x values are introduced. These are chosen to only contain one decimal place in order not to overcomplicate the task. The numbers are chosen so that students who are uncomfortable to end up with a rational solution, where $0 < x < 1$, are incentivised to carry out inverse operations, resulting in the student answering $x = 2$, instead of $x = 0.5$. In other cases, a student may multiply instead of divide, to make the equation make sense.

5.2.4. Type 5: $x/b = c$

This type of equation follows a similar pattern to type 4 (see Figure 1). Technically, the number b could be considered the fraction $\frac{1}{b}$ in type 4, in which case we would end up with type 5. However, one of the main criteria that restricts our design was the decision to use integers only, and in some cases, rather simple decimal numbers. Table 6 exemplifies this.

Table 6. Examples of type 5 equations and variations.

Original	Variation 1	Variation 2	Variation 3	Variation 4	Variation 5
$\frac{x}{3} = 9$ $\frac{x}{2} = 2$	$\frac{x}{9} = 3$	$\frac{x}{2} = -9$	$\frac{x}{-3} = 9$	$\frac{x}{-3} = -9$	$\frac{x}{4} = 3.4$ $\frac{x}{-4} = 3.4$

The original variation also uses number pairs (a, c), where a is a factor of c. From the literature we know that some students may be tempted to divide c by a instead of multiplying. Variation 1 features equations whose number pairs are exchanged. This may be expected to result in a 'crisis' [53] for those students who divide instead of multiply. Variations 2, 3 and 4 use combinations of negative numbers, as do the preceding types. Variation 5 uses rational numbers for the c value, and integers for the a value. The value of a remains an integer, since multiplication with rational numbers is beyond the scope of this study (and design).

5.2.5. Type 6: $a/x = c$

This form of equation is the third and final version of the single operation variant that centres on multiplication (see Figure 1). What is different about this variation is that technically, these must be considered two-step equations. In these equations, the unknown is positioned in the denominator, which makes a tremendous difference in complexity, compared to the previous type (type 5). That said, some straightforward strategies may be easily applied to solving these problems. The 'guess-and-check' or a knowing strategy may be very efficient here, if the students are able to read and understand what the problem requires of them. To solve these formally, one would have to multiply each side by x before dividing each side by c, which is known to be a demanding procedure for some students. Table 7 provides examples.

Table 7. Examples of type 6 equations and variations.

Original	Variation 1	Variation 2	Variation 3	Variation 4
$\frac{9}{x} = 3$ $\frac{2}{x} = 2$	$\frac{9}{x} = -3$	$-\frac{9}{x} = 3$	$-\frac{9}{x} = -3$	$\frac{6.6}{x} = 2$ $\frac{4.5}{x} = 9$ $-\frac{6.6}{x} = 2$

The original variation actually uses the numbers in pairs, as they were used in the previous type (type 5). This makes future comparisons possible. The intent is to observe the extent to which type 5 and type 6 equations may generate the same, or similar, wrong answers. Variation 1 is not simply a switching of a and c (as seen in type 5), since this would result in a solution that exceeds scope of this study. Hence, variations 1, 2 and 3 have negative numbers for a and/or c, which gives negative solutions for variations 1 and 2, and a positive solution for variation 3. Variation 4 has rational numbers for the a values, and therefore has rational solutions. The c value is kept as a natural number in order to determine whether this suggests to students that the solution is natural.

5.2.6. Type 7: $ax + b = c$ and $ax - b = c$

This type of equation is the first to include two operations on the left hand side of the equals sign, and which introduces a third constant to the task design. To many teachers and researchers, these equations may resemble classic arithmetical equations [12]. However, some level of distinct abstraction is possible. First, with these equations we must move further away from the application of the 'guess-and-check' strategy, and toward more refined solution strategies, which, according to the literature, are likely to be required. These equations must also be considered two-step equations, with $a \notin \{0,1\}$. To some extent, equation type 3 ($b - x = c$, cf. Figure 1) does require two steps for its solution, but does not require two operations on the left side, and is arguably easier to solve. The reason for using the minus sign between the terms ax and b for the categorisation is the immediate jump in the level of abstraction involved (e.g., [14]). The design involves manipulating the values of a and c accordingly, to identify the subtypes. See Table 8 for examples.

Table 8. Examples of type 7 equations and variations.

Original	Variation 1	Variation 2	Variation 3	Variation 4	Variation 5
$2x + 7 = 21$ $2x - 7 = 21$	$8x + 16 = 8$	$8x + 10 = 30$ $8x + 30 = 10$	$-7x - 9 = 33$ $-2x - 5 =$ -3	$4x + 12 =$ -10 $-6x + 2 =$ -13	$10 = 2x + 6$

The original variation is the most straightforward. The a, b and c are natural numbers with values less than 30, which is intended to ensure positive natural number solutions ($c > b$). In contrast, we have an identical task design idea, but with the b value inverted. The point here is to develop a task where a is a true divisor in $c - b$, $b - c$, $c + b$ and $b + c$. This form ensures that the equations are interesting from a data-analysis perspective, and may encapsulate the students' difficulties with this setup or these types of equations.

Variation 1 is similar to the first subtype with respect to the a, b and c values. However, now $b > c$, to generate negative integer solutions. Again, a is kept as a true divisor in $c - b$, $b - c$, $c + b$ and $b + c$. Variation 2 has a positive rational number solution. The reason for restricting the design to generate positive solutions is to enable future analyses of students' difficulties. If the tasks have both positive and negative solutions, this complicates the distinction between difficulties related to negatives and difficulties related to rational numbers. The solutions are limited to $p/2$, where p is a natural number. The reason for restricting the design in this way is to avoid over complication. The tasks are intentionally designed to take advantage of students' willingness to alter operations (for the sake of sense-making). In these cases, some students may be tempted to add 30 to 10 or 10 to 30, to reach an integer solution. Variations 3 and 4 use negative a and c integers. Variation

3 keeps the integer solutions, whereas variation 4 advances to rational number solutions ($p/2$, where $p \in Z$). Variation 5 introduces a reversed version of the equation, similar to those in types 1 and 4.

5.2.7. Type 8: $a(x+b) = c$

This type of equation is the first of three versions of the classic arithmetical equation $ax + b = c$ (type 7). In many cases, this variation of $ax + b = c$ may be regarded as easy to solve. However, some students may be challenged by the necessity of working with calculations with brackets. Technically, the brackets are easily managed in this type of task, since the equations may be rewritten by dividing by a on each side. Predictably, many students will begin by expanding the brackets, and many may have difficulty with this expansion, and end up with an equation that is not equivalent to the initial equation [6]. Presumably, some students may observe the factorisation of c into a and $(b + x)$. With this train of thought, a 'working backwards' or 'undoing' strategy would be suitable for these equations. This type of equation provides a direct contrast to the type 7 ($ax + b = c$) equation. Arguably, for this variation it may be difficult to apply a 'guess-and-check' strategy, when compared to type 7. For examples, see Table 9.

Table 9. Examples of type 8 equations and variations.

Original	Variation 1	Variation 2	Variation 3	Variation 4	Variation 5
$2(x+4) = 12$	$2(x+12) = 20$	$3(x-5) = 6$	$2(x+4) = 8$	$2(x+4) = 13$	$-2(x+3) = -10$ $-6(x+8) = -12$

The original variation consists of 'pretty' versions of these equations, with natural values for a, b and c. The equations have natural number solutions because a is chosen as a factor of c. Variation 1 also uses natural numbers for a, b and c, and a as a factor of c. However, in these instances, a times b is greater than c, which results in negative solutions. Variation 2 includes subtraction in the brackets: in other words, a negative b value and positive solutions. Variation 3 has *0* as a solution, because a times b equals c. Variation 4 has rational numbers as solutions. The solutions are limited to $p/2$, where p is a natural number. Variation 5 uses a negative a and c, with both positive and negative integer solutions.

5.2.8. Type 9: $a/x + b = c$

This type of equation is the second variation of type 7 ($ax + b = c$), but also has many technical similarities to type 6 ($b/x = c$). Formally, solving these equations requires three steps. The reader may agree that the likelihood of successfully applying a 'guess-and-check' or a 'knowing' strategy at this stage is unlikely. To some extent a 'cover up' or a 'working backwards' strategy may work, and the task is rather easily transformed into a type 6 equation. Table 10 provides examples.

Table 10. Examples of type 9 equations and variations.

Original	Variation 1	Variation 2	Variation 3	Variation 4
$\frac{3}{x} + 2 = 5$ $\frac{6}{x} + 3 = 5$	$\frac{6}{x} + 5 = 3$	$\frac{-4}{x} + 4 = 2$ $\frac{-4}{x} + 2 = 4$	$\frac{9}{x} - 2 = 1$ $\frac{9}{x} - 5 = -8$	$\frac{7.5}{x} + 5.5 = 8.5$

The original equation uses natural numbers and natural number solutions, with $b < c$. In variation 1, $c < b$, to yield negative integer solutions. In variation 2, a has a negative integer value, with the added variation of $c < b$ and $b < c$ (b and c are exchanged). Both positive and negative solutions may result, because of the relationship between c and b. Variation 3 includes a subtraction or a negative b value. Both positive and negative solutions may result, owing to positive and negative c values. Variation 4 has rational a, b and c values. Again, the values are limited to $p/2$, where p is a natural number.

5.2.9. Type 10: $a/(x+b) = c$

This final variation of type 7 equations ($ax + b = c$) may be considered the most difficult or complex of the arithmetical equations. However, to develop a more advanced three/four-step equation, we designed this version of type 7, where the denominator includes an operation. In these tasks, too, the students must work with brackets to find a formal solution (cf. type 8). A 'guess-and-check' strategy may not be out of the question, but a 'cover up' strategy may be more fruitful in this situation. A 'formal strategy' would involve multiplying with the denominator on each side, yielding a type 7 ($ax + b = c$) equation. Presumably, this would be the strategy of choice for students who are able to solve these equations. See Table 11 for examples.

Table 11. Examples of type 10 equations and variations.

Original	Variation 1	Variation 2	Variation 3	Variation 4	Variation 5
$\frac{6}{x+1} = 3$ $\frac{15}{x+3} = 3$	$\frac{15}{x-2} = 3$	$\frac{5}{x+3} = 5$ $\frac{12}{x+8} = 3$	$\frac{12}{x-7} = -4$	$\frac{-15}{x-6} = 15$ $\frac{-15}{x-1} = 3$	$\frac{19}{x+7} = 2$ $\frac{19}{x+12} = 2$

The original variation of this equation uses natural numbers and natural number solutions. Variation 1 involves subtraction from the unknown in the denominator, but still with natural number solutions. Variation 2 consists of addition in the denominator (as in the original variation), but with negative integer solutions. Variation 3 involves subtraction in the denominator, but now with negative integer solutions, because of a negative c value. Variation 4 uses a negative value of a, with both positive and negative solutions. Variation 5 has rational number solutions limited to $p/2$, where p is a natural number.

5.3. Algebraic Equations

This subsection discusses algebraic equations [12]. These equations have the unknown in two terms, on either side of the equals sign. In the literature, it is evident that difficulties similar to those described for arithmetical equations may be observed for the algebraic. However, algebraic equations present the added difficulty of requiring operations with an unknown [12,14]. 'Simplified algebraic equation' is as mentioned the term we use for equations that have the form, $ax + b = cx + d$, where a, b, c and d are known, and x is the unknown for which we solve. Simplified algebraic equations involve other aspects related to students' difficulties, which separates them from the simplified arithmetical equations. Algebraic equations have multiple terms to the right of the equals sign. This addresses the acceptance of the problem of a lack of closure [32]. With arithmetical equations, a student may apply an arithmetic understanding of the equals sign, and to a great extent, view the equation as a simple calculation. A student cannot apply this view to algebraic equations.

Type 11: $ax + b = cx + d$

Before presenting the variations of algebraic equations, we must establish a perhaps not-so-obvious—or underappreciated—distinction. When accustomed to solving arithmetical equations, many beginning equation-solvers predictably wish to isolate the unknown to the left hand of the equals sign. The solution to an equation is usually presented or written as $x = p$, where x is the unknown. Bearing this in mind, there emerges a clear distinction between algebraic equations in which c is larger than a, and equations where a is larger than c. Of course, this is considered from a strategic perspective. Nonetheless, the variations begin with this distinction. However, we will limit the scope of the task design to simplified algebraic equations, which comprise algebraic equations of the form, $ax + b = cx + d$, where a, b, c and d are constants, and x represents the unknown. It is important to remember that if students are not applying a 'guessing' or a 'knowing' strategy to these equations, they must operate on the unknown. This observation makes it important to design tasks to address the role of the coefficients of the unknowns present. Conventions dictate that we do not have a present coefficient when it is equal to *1*. Errors such as conjoinment (e.g., [11,19]) are an additional factor in the design. We consider the two general cases of variation in the simplified algebraic equations, where $a > c$ and $a < c$. Examples are given in Tables 12 and 13.

Table 12. Examples of type 11 equations and variations with $a > c$.

Original	Variation 1	Variation 2	Variation 3	Variation 4
$2x + 1 = x + 8$	$2x + 8 = x + 1$	$5x + 13 =$ $-5x + 33$	$2x + 5 = 1x + 5$	$3x + 8 = x + 3$

Table 13. Examples of type 11 equations and variations with $a < c$.

Original	Variation 1	Variation 2	Variation 3	Variation 4
$x + 8 = 2x + 1$	$x + 1 = 2x + 8$	$-5x + 33 =$ $5x + 13$	$1x + 5 = 2x + 5$	$x + 3 = 3x + 8$

The original variation has a natural number solution, and has only natural number entries. This variation also uses c values between 1 and 20 (bear in mind that this type use a values greater than c values). In variation 1, the b and d values are exchanged to obtain a negative solution. Variation 2 uses c as an additive inverse of a. This addresses the question of inverting subtraction with subtraction, or inverting addition with addition. If this is the case, some students may find that the unknown has 'disappeared', thus creating a scenario in which any number may be a solution. The literature indicates that if this happens, some students may become confused, and conclude that $x = 0$. Variation 3 uses identical b and d values, which gives 0 as the solution. Variation 4 uses rational numbers as solutions.

These variations are simply reversed variations of the equations in Table 12. These are presented to investigate the extent to which equations where it is not the ideal strategy to collect the unknowns on the left side of the equality sign are more difficult than those in which it is. Ultimately, these equations are identical to the equations in Table 12.

5.4. Different Branches of Variations of Linear Equations

In this subsection, we touch on some of the slight variations in the concept of the linear equation. In all the above-mentioned varieties of linear equation we used the letter x to represent the unknown. However, the literature [44] mentions that some students encounter difficulties when the name or the label of the unknown is an unfamiliar letter. In Denmark, the letter most commonly used to represent an unknown in a linear equation is x. Therefore, we have chosen to present variations on some of the above-mentioned types of equations with the letter x exchanged for a different one. For this we chose, types 1, 2, 4 and 7. We chose these to capture a broad variety of tasks in the section with arithmetical equations. We chose to not include an algebraic version, since we already made a comparison of arithmetic and algebraic equations, and we saw no need to present another, similar, comparison. Tables 14–16 present the specific equations we use in the dynamic online environment.

Table 14. Examples of type 1 and type 2 equation variations with n as the unknown.

Original	Variation 1	Variation 2	Variation 3	Variation 4	Variation 5
$4 + n = 7$	$7 = n + 4$	$7 + n = 4$	$4 + n = -7$	$7 + n = -4$	$n + 4 = 7$
$n - 3 = 5$	$5 = n - 3$	$n - 5 = 3$	$n - 3 = -5$	$n - 5 = -3$	$-5 + n = 3$

Table 15. Examples of type 4 equation variations with n as the unknown.

Original	Variation 1	Variation 2	Variation 3	Variation 4	Variation 5
$3n = 15$ $8n = 8$	$3n = -15$	$-3n = 15$	$-3n = -15$	$15 = 3n$	$28n = 14$ $2n = 9$

Table 16. Examples of type 7 equation variations with n as the unknown.

Original	Variation 1	Variation 2	Variation 3	Variation 4
$2n+7=21$ $2n-7=21$	$8n+16=8$	$8n+10=30$ $8n+30=10$	$-7n-9=33$ $-2n-5=-3$	$4n+12=-10$ $-6n+2=-13$

We now briefly address the question raised by Vlassis [14] regarding Filloy and Rojano's [12] categorisation of linear equations. Vlassis [14] pointed out that the original categorisation was based on the 'didactic cut' [12], which some consider problematic. This has to do with abstract arithmetical equations [14]. Although many of these abstract arithmetical equations are covered by types 1 through 10, we intentionally excluded what we call *non-simplified arithmetical equations* (subset of pre-algebraic equations [14]). By non-simplified we mean that the unknown is represented in multiple terms, but on only one side of the equals sign. To complete our design, we included variations of the following equation template, $ax + b + cx = d$, as shown in Table 17.

Table 17. Examples of non-simplified equations and variations based on $ax+b+cx=d$ [14].

Original	Variation 1	Variation 2	Variation 3	Variation 4
$x+6+2x=15$	$4x-3+2x=15$	$4x+3-2x=$ $134x-3-2x=11$	$2x+3-4x=13$	$-2x+3-4x=21$

The original variation uses positive a and c values, and a natural number for $d-b$. This results in natural number solutions. The idea is to measure how these non-simplified arithmetical equations perform when compared to simplified arithmetical equations. The number of variations is restricted for this particular reason. The intension is not to present every variation of a simplified arithmetical equation (types 1–7) transformed into an equivalent non-simplified version. Variation 2 uses a negative integer value for b, still with natural number solutions. Variation 3 reverses the a and c values. This results in $a + c$ being a negative integer, and in negative integer solutions. Variation 4 uses negative integer values for a and for c. This still yields negative integer solutions.

5.5. Limitations and Opt-Outs

Some initial limitations, which we discovered in the early phase of the design process, were those set by the format of the dynamic online environment. As mentioned, we are limited to input fields handling only rational numbers (as decimal numbers) and multiple-choice. The intention, regarding the future data analysis, was thus to create a large amount of distinguishable linear equations in order to make a successful attempt of categorizing students as equation-solvers. Other approaches to diagnostics require in-depth analysis of students' mathematical behaviour through conversations, interviews, and structured situations [3].

Due to the limitations of the input methods in the dynamic environment, we have also opted out designing linear equations with infinitely many or no solutions. If we were to implement such equations in the online environment, we would have to use the multiple choice input method. As previously explained, this was deemed undesirable.

6. Discussion

The question we initially set out to answer was, '*How may research on lower secondary school students' difficulties with linear equations inform task design in a dynamic online environment with the possibility to promote/support students' personalised learning?*' As explained, we approached this question by applying a two-tier methodology: the first tier involved reviewing the existing literature on students' difficulties with solving linear equations; the second tier involved applying theoretical constructs from task design, in particular variation theory, to developing tasks that promote students' personalised learning. As is evident from the final design presented above, although these two tiers initially ran along separate tracks, they eventually became rather entangled. Our discussion below examines this.

First, we would like to discuss why we chose to address the aspects of the literature that are presented in Section 3. When we carried out the snowballing portion of the hermeneutic review process, a picture of the general categories of the literature began to take shape, including the questions that this research literature addresses has dealt with. This prompted a deeper dive into the general categories of the literature, while we kept in mind the future implementation linear equation tasks in a dynamic online environment. As we mentioned, we knew that we would end up with a design that consisted of a number of easily accessible tasks, not being one combined diagnostic task [52]. We wished to determine whether quite ordinary training tasks may be designed and implemented to address the difficulties that lower secondary school students encounter when solving equations. It was also important to us to keep our focus on the relevant age group at all times, while snowballing the literature.

As lower secondary school students are our focus, the difficulties mentioned in the literature gave rise to the design of assignments for this particular group of students. In Danish lower secondary schools, algebraic equations are not a dominant topic. Many of the difficulties that the literature describes as surrounding the concept of the equation may be addressed by using arithmetical equations. We have thus chosen to put a greater emphasis on designing such equations. This was done mainly because of the target group. In lower secondary school, many students struggle with the concept of equations [6]. We also know from the literature that students' difficulties arise when operations have to be done on the unknown or the unknown is present on both sides of the equal sign [6,12–14]. However, several of the remaining difficulties (specifically issues regarding the presence of negative numbers) that students experience during equations solving can arise when solving arithmetical equations. For this reason, we have chosen to focus the task design more towards arithmetical equations and negative numbers. However, it is clear that the difficulties that may arise because of the presence of the unknown in several terms of an equation may be addressed only via a non-simplified arithmetical equation or an algebraic equation. This conclusion underlies the significant difference in the number of types of equations presented in the general division into arithmetical equations and algebraic equations (cf. Figure 1). This conclusion, together with the difficulties described in the literature, also underlies the way the various types of equations are designed. This is expressed in the learning tree of arithmetical equations (Figure 1) and the single type of algebraic equations (type 11). Thus, the difficulties described in the literature influenced the design process in two ways. First, the general categories of documented difficulties of the target group gave rise to the types of equation that are deemed relevant. Second, the theory of variation used on the more nuanced documented difficulties within each category gave rise to the varieties that are considered relevant to each type of equation.

Next, we discuss the mathematical elements present in the general equation types and their varieties. First of all, the opt-outs of mathematical content related to students' documented difficulties, were made during the development of the tasks. Fractional coefficients have not been prioritised here, since they are not regarded as having a direct impact on students' understanding of the concept of the equation. Surely, the concept of the fraction is difficult. However, we say that the concept of the fraction is difficult for lower secondary school students, regardless of whether or not a fraction appears in an equation. That said, there are a few fractions in equation types 5, 6, 9 and 10. However, these fractions can be replaced with another syntax indicating division, meaning the fractions in the task design are purely operational. When examining the literature (e.g., [6]), we find that fractions do in fact present a source of great difficulty for younger students, who are doing algebra and arithmetic. We have not included fractions (as coefficients, terms or solutions) for two reasons. The first being that the online environment does not support fractions as an input in the input field, which rules fractions as solutions out. Fractions as coefficients and terms have not been a part of the design due to the second reason, which can be stated as a question: Fractions do arguably make algebra and arithmetic more difficult, but do fractions make equations more difficult? Knowing that a student cannot solve equations with fractions for *a,b,c* or *x*, does not mean that the student does not know how to solve equations, rather that it means that the student has difficulties working with fractions. However, we wanted to make it possible to evaluate as many varied situations as possible in which the unknown must be treated differently. In type 5, ($x/b = c$), we decided that a student should be able to handle the fraction as a process, i.e., the process it may be seen as, rather than using other mathematical symbols to indicate division. On the other hand, this also clarifies the scope of the findings that address difficulties with the concept of the equation, and solving equations. For us, the extent of the difficulties surrounding negative numbers, or the concept of numbers in general, is now much clearer; this also applies to working with the concept of the equation. It may be difficult to make a concrete comparison of the difficulties. Nevertheless, difficulties surrounding negative numbers—as solutions, coefficients and constant terms—may be observed at every step of a learning path. However, we have strongly emphasised negative numbers as solutions, as constant terms and as coefficients. With regard to the literature on difficulties with the concept of the equation, it is striking how much emphasis is actually placed on negative numbers. Lower secondary school students form our target group, and most of these students may be considered inexperienced equation solvers. Several research studies place great emphasis on the fact that negative numbers and the use of the negative generally becomes a focal point when teaching the concept of the equation [6,14,18]. Therefore, negative numbers became a pervasive element of our designs. To generate a negative solution to a standard arithmetical equation ($ax + b = c$), either b must be greater than c, or a minus sign must be present in some way. If we factor this into our design, we may see a difference in the answers to tasks, where negative solutions emerge with or without the presence of the minus sign.

The second category of difficulty in solving equations concerns the equals sign. Much of what is emphasised in the literature is not necessarily easy to transfer to a design that must consist of simple equation-solving tasks. Many might argue that if it is to be possible to determine whether a student has a good understanding of, or has difficulty in interpreting the equals sign in a given situation. In this case, some completely different tasks may be needed. Much of what is described regarding the equals sign has to do with relation versus process. Based on our task design, it would be inaccurate to claim that we may determine whether or not a student has a strong grasp of the equals sign. However, what it has been possible to apply to the design is the question of the 'acceptance of lack of closure' [32]. Here, we have been able to develop variations of individual equation types, in order to investigate whether or not everyone who can solve the equation $ax = b$ can also solve the equation $b = ax$. We have done this with equation types 1, 4 and 7. If a student has difficulty with 'the acceptance of lack of closure', algebraic equations may easily present an insurmountable challenge. A rigid arithmetic interpretation of the equals sign may lead to omission of the d term in algebraic equations such as $ax + b = cx + d$ [23,27].

Another important question related to the task design concerns the deliberate omission of the multiplication sign. In Denmark, it is customary to indicate the multiplication of two numbers as $a \cdot b$. This is not the case in all countries: for example an '$\times$' is used in some English-speaking countries. In general, in Danish lower secondary schools, arithmetical equations are presented in the form $ax + b = c$, with the multiplication sign omitted. We have chosen to design and present our tasks in the same way, with the multiplication sign omitted. This is also partly influenced by the format of the dynamic online environment in question, *matematikfessor.dk*.

As previously mentioned, we wanted to make sure that we were likely to have a clear way of separating the linear equations, or types of linear equations included in our task design. We have also mentioned that we designed a reasonably comprehensive arsenal of tasks for beginning or early intermediate equation solvers (e.g., Danish lower secondary school students in the 7th year). The task design must form a learning path that serves as a tool for detecting students with difficulties related to linear equations. We suggest that the tree (Figure 1) that describes the relationship of the types of arithmetical equations may give rise to a learning path by traversing its branches. When a student has to work with arithmetical equations, the underlying structure of this tree may clarify the connections between difficulties and progress. This is not to say that every branch of the tree represents difficulties that cannot be found in other branches of the tree. The tree gives teachers and students the opportunity to look at the difficulties that arise while students progress along the mathematics learning path, and therefore they should try to take a step back to identify the cause of their difficulties. It may be said that students may have deeper mathematical difficulties that are not directly related to equation-solving, but related to embedded elements such as the concept of the number, the equals sign, and so forth, and their errors when solving equations are to be considered symptoms of something more deeply rooted [3,22].

With regard to the goal of personalised learning, the development of comparable tasks became an integral part of the task design. At best, a large data set of students' answers to these exercises should reveal patterns of students as equation solvers'. When we say 'comparable', we mean that many of the task variations have been constructed in such a way that e.g., simply swapping two values within an equation creates a completely different solution. As mentioned, the possibility of substantial data extraction is one of the obvious benefits of including the tasks of the 11 types of linear equations in a dynamic online environment. Hence, selecting the parts of the literature that described students' difficulties with the concept of linear equation, and which enabled us to form comparable tasks, became our mission. It must be said that when the literature on students' difficulties with equation-solving is studied as far back as five decades, it may be difficult to determine how time has left its mark on the implementation of the curriculum and the expectations of today's student, in terms of equation-solving. From previous studies [22], we know that in Denmark, upper secondary school students have significant difficulties with the concept of the equation and with equation-solving. Therefore, this initiative may provide teachers with an opportunity to become more familiar with, and thereby focus more on, the concept of the equation in lower secondary schools.

The guidelines presented by Watson and Mason [52] had a major impact on our overarching, initial design requirements. It helped us make clear which elements and sub-concepts of the concept of the equation we wanted address. Additionally, we made it possible for mathematical concepts that may cause students difficulties to come into play in the equations. From the perspective of the individual task, variation theory [51] and the element of the crisis [53] played major roles. If the equations of types 1 through 11 are seen in the context of problem difficulty [50], then an increase in difficulty may be noted, not only in the progression of types, but also in the variations of each type. The complexity increases through this progression, as the number of terms, operations on the unknown and the contexts in which the unknown is included, increase. Unrecognizability is addressed through the many variations, where apparently similar equations may require very different solutions (negative and rational numbers). The technical demands are also increased through the progression of the types. The unknown is part of the more advanced constellations of the ten arithmetical equation types (1–10), and also in the leap to algebraic equations (type 11), where the unknown occurs on both sides of the equals sign. It was not possible to address the question of the student's autonomy [52] when solving problems. This is solely because the tasks need to be implemented in a dynamic online environment. Many of the tasks in the design also seek to address some of the extreme examples of equations that may easily be overlooked or omitted during ordinary classroom teaching. Such extreme cases may lead to crises [4] that ensure that as many as possible of the categories and subcategories of difficulty identified have the opportunity to come into play.

It should be noted that though we have focused on equation-solving at Danish upper secondary schools, the research shows that students' mathematical difficulties, including difficulties with equation-solving, are not determined by national borders [3]; if not universal, they are at least international in nature. A remark may also be made with respect to the task design being affected by a specific dynamic online environment, which is *matematikfessor.dk*. Given that the mathematics difficulties we have addressed are universal, the ways in which they need to be addressed are bound to be similar on various online platforms.

7. Conclusions

We set out to address the question of how five decades of research into students' difficulties with equation-solving may inform the design of the now-widespread, dynamic online environments to which many lower secondary school students turn for everyday mathematics instruction, and how the designs of these potentially may promote and support students' personalised learning. Our approach to this research question has been to present a concrete design of eleven types of equations, including carefully selected variants of these types, to illustrate possible personalised learning paths as the branches of a tree construct that comprises these equation types. By making students and teachers aware of existing difficulties, and the positions of these in the tree structure, teachers may have an additionally unique opportunity to help students to overcome these difficulties as part of their teaching. The future role of the dynamic online environment will be to not only identify students, who have difficulty solving linear equations, and to diagnose the nature of their difficulties, but eventually to also intervene, although preferably in some sort of collaboration with the student's mathematics teacher. Once a student's difficulties are diagnosed, the dynamic online environment may have the student explore all the aspects of these, by carefully traversing the variations in the tree structure. However, for this to happen, one must begin with the concrete design of types of equations. We have attempted to present such a design by drawing on previous research on task design, both with and without digital technology, and particularly by applying the accumulated research on students' difficulties with linear equation-solving where it belongs: to the students.

Author Contributions: Conceptualization, M.E. and U.T.J.; methodology, M.E. and U.T.J.; software, M.E. and U.T.J.; formal analysis, M.E. and U.T.J.; investigation, M.E. and U.T.J.; data curation, M.E. and U.T.J.; writing—original draft preparation, M.E. and U.T.J.; writing—review and editing, M.E. and U.T.J.; supervision, U.T.J.; funding acquisition, M.E. All authors have read and agreed to the published version of the manuscript.

Funding: The study presented in this paper is funded by *Innovation Fund Denmark* by grant no. 8053-00042A.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: Thank you to the anonymous reviewers for constructive comments and recommendations and to Michaela Scioscia for English editing.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Table A1. References That Formed the Basis for the Literature Snowballing.

Title	Journal	Authors	Year
Misconceptions and learning difficulties of captured students enrolled in development mathematics courses.	Ohio Journal of School Mathematics	Ahuja Om; Najafi M;	2003
Middle School Students' Conceptual Understanding of Equations: Evidence from Writing Story Problems	International Journal of Educational Psychology	Alibali Martha W; Stephens Ana C; Brown Alayna N; Kao Yvonne S; Nathan Mitchell J;	2014
A hypergraph-based framework for intelligent tutoring of algebraic reasoning.		Arevalillo-Herráez Miguel; Arnau David;	2013
Specularity in algebra.	For the Learning of Mathematics	Asghari Amir;	2012
An interactive algebra course with formalised proofs and definitions.		Asperti Andrea; Geuvers Herman; Loeb Iris; Mamane Lionel Elie; Sacerdoti Coen; Claudio;	2006
Middle school mathematics teachers' knowledge of students' understanding of core algebraic concepts: equal sign and variable.	Mathematical Thinking and Learning	Asquith Pamela; Stephens Ana C; Knuth Eric J; Alibali Martha W;	2007
Struggling with variables, parameters, and indeterminate objects or how to go insane in mathematics.		Bardini Caroline; Radford Luis; Sabena Cristina;	2005
Modes of algebraic communication: moving from spreadsheets to standard notation.	For the Learning of Mathematics	Bills Liz; Ainley Janet; Wilson Kirsty;	2006
Making Sense of Integer Arithmetic: The Effect of Using Virtual Manipulatives on Students' Representational Fluency	Journal of Computers in Mathematics and Science Teaching	Bolyard Johnna; Moyer-Packenham Patricia;	2012
Misconceptions and learning algebra.		Booth Julie L; McGinn Kelly M; Barbieri Christina; Young Laura K;	2017
Understanding Problem-Solving Errors by Students with Learning Disabilities in Standards-Based and Traditional Curricula	Learning Disabilities: A Multidisciplinary Journal	Bouck Emily C; Bouck Mary K; Joshi Gauri S; Johnson Linley;	2016
Children Learn Spurious Associations in Their Math Textbooks: Examples from Fraction Arithmetic	Grantee Submission	Braithwaite David W; Siegler Robert S;	2018
Basic algebra problems through the calculator based computational approach.	Acta Didactica Universitatis Comenianae. Mathematics	Brody Jozef; Rosenfield Steven; Lytle Pat;	1993
Eighth Grade Students' Representations of Linear Equations Based on a Cups and Tiles Model	Educational Studies in Mathematics	Caglayan Gunhan; Olive John;	2010
Different Grade Students' Use and Interpretation of Literal Symbols	Educational Sciences: Theory and Practice	Celik Derya; Gunes Gonul;	2013
Individual differences in the mental representation of term rewriting.		Cohors-Fresenborg Elmar;	2002
Using the number line to investigate the solving of linear equations.	For the Learning of Mathematics	Dickinson Paul; Eade Frank;	2004

Table A1. *Cont.*

Title	Journal	Authors	Year
Helping Students with Mathematics Difficulties Understand Ratios and Proportions	Teaching Exceptional Children	Dougherty Barbara; Bryant Diane Pedrotty; Bryant Brian R; Shin Mikyung;	2016
Is Algebra Really Difficult for All Students?	Acta Didactica Napocensia	Egodawatte Gunawardena;	2009
Transition from arithmetic to algebra in primary school education.	Teaching Mathematics and Computer Science	Fülöp Zsolt;	2015
2x minus x equals 2.	The New Zealand Mathematics Magazine	Gage Jenny;	2002
Basic Arithmetical Skills of Students with Learning Disabilities in the Secondary Special Schools: An Exploratory Study Covering Fifth to Ninth Grade	Frontline Learning Research	Gebhardt Markus; Zehner Fabian; Hessels Marco G. P;	2014
Pre-Service Middle School Mathematics Teachers' Understanding of Students' Knowledge: Location of Decimal Numbers on a Number Line	International Journal of Education in Mathematics, Science and Technology	Girit Dilek; Akyuz Didem;	2016
Getting to grips with 'equals'—A balancing act.	Equals [electronic only]	Haseler Margaret;	2010
The space between the unknown and a variable.		Hewitt Dave;	2014
Construction of an Online Learning System for Decimal Numbers through the Use of Cognitive Conflict Strategy	Computers and Education	Huang Tzu-Hua; Liu Yuan-Chen; Shiu Chia-Ya;	2008
How close do we need to be?	Mathematics Teaching	Hughes Mervyn;	2014
Some issues in assessing proceptual understanding.		Hunter M; Monaghan J;	1996
Preservice Teachers' Knowledge of Students' Cognitive Processes about the Division of Fractions	Hacettepe University Journal of Education	Isiksal Mine; Cakiroglu Erdinc;	2008
Algebra homework. A sandwich!	Mathematics Teacher	Jackson D Bruce;	2014
The Contribution of Domain-Specific Knowledge in Predicting Students' Proportional Word Problem Solving Performance	Society for Research on Educational Effectiveness	Jitendra Asha K; Lein Amy E; Star Jon R; Dupuis Danielle N;	2013
Exploring the meaning of letters.	Mathematics Teaching	Jones Martin;	2012
Difficulties in Initial Algebra Learning in Indonesia	Mathematics Education Research Journal	Jupri Al; Drijvers Paul; van den Heuvel-Panhuizen; Marja;	2014
Early Developmental Trajectories toward Concepts of Rational Numbers	Cognition and Instruction	Kainulainen Mikko; McMullen Jake; Lehtinen Erno;	2017
The study on variable substitution in learning mathematics.	Far East Journal of Mathematical Education	Kang Jeong Gi;	2013
A new curriculum for structural understanding of algebra.	Journal of the Korean Society of Mathematical Education. Series D	Kirshner David;	2006
What Do Error Patterns Tell Us about Hong Kong Chinese and Australian Students' Understanding of Decimal Numbers?	International Journal for Mathematics Teaching and Learning	Lai Mun Yee; Murray Sara;	2014
Struggling to disentangle the associative and commutative properties.	For the Learning of Mathematics	Larsen Sean;	2010

Table A1. *Cont.*

Title	Journal	Authors	Year
Sources of differences in children's understandings of mathematical equality: Comparative analysis of teacher guides and student texts in China and the United States.	Cognition and Instruction	Li Xiaobao; Ding Meixia; Capraro Mary Margaret; Capraro Robert M;	2008
The usefulness of an intensive diagnostic test.	Pythagoras (Pretoria)	Liebenberg Rolene;	1998
An Error Analysis of Form 2 (Grade 7) Students in Simplifying Algebraic Expressions: A Descriptive Study	Electronic Journal of Research in Educational Psychology	Lim Kok Seng;	2010
Conceptual maps and equations: What is the meaning of this?	Mediterranean Journal for Research in Mathematics Education	Lima Rosana Nogueira de;	2008
Concept Development of Decimals in Chinese Elementary Students: A Conceptual Change Approach	School Science and Mathematics	Liu Ru-De; Ding Yi; Zong Min; Zhang Dake;	2014
Proficiency in the Multiplicative Conceptual Field: Using Rasch Measurement to Identify Levels of Competence	African Journal of Research in Mathematics, Science and Technology Education	Long Caroline; Dunne Tim; Craig Tracy S;	2010
Why Is Learning Fraction and Decimal Arithmetic so Difficult?	Grantee Submission	Lortie-Forgues Hugues; Tian Jing; Siegler Robert S;	2015
"But What about the Oneths?" A Year 7 Student's Misconception about Decimal Place Value	Australian Mathematics Teacher	MacDonald Amy;	2008
From research on student difficulties in using the properties of functions while solving equations and inequalities.		Major Joanna; Powzka Zbigniew;	2009
The interweaving of arithmetic and algebra: some questions about syntactic and structural aspects and their teaching and learning.		Malara Nicolina A; Iaderosa Rosa;	1999
Unknown or 'thing that varies'? The implicative statistic analysis and the factorial analysis of the correspondences in a research in mathematics education.	Acta Didactica Universitatis Comenianae. Mathematics	Malisani Elsa; Spagnolo Filippo;	2005
Teaching structure in algebra.	Mathematics Teacher	Merlin Ethan M;	2013
Using Habermas' theory of rationality to gain insight into students' understanding of algebraic language.		Morselli Francesca; Boero Paolo;	2011
An Examination of the Ways that Students with Learning Disabilities Solve Fraction Computation Problems	Elementary School Journal	Newton Kristie J; Willard Catherine; Teufel Christopher;	2014
Pilot Study on Algebra Learning among Junior Secondary Students	International Journal of Mathematical Education in Science and Technology	Poon Kin-Keung; Leung Chi-Keung;	2010
Assessing Knowledge of Mathematical Equivalence: A Construct-Modelling Approach	Journal of Educational Psychology	Rittle-Johnson Bethany; Matthews Percival G; Taylor Roger S; McEldoon Katherine L;	2011
Designing innovative learning activities to face difficulties in algebra of dyscalculic students: exploiting the functionalities of AlNuSet.		Robotti Elisabetta;	2017
The difficulties in the search for solutions of functional inequalities.	Mathematics Competitions	Samovol Peter; Zhuravlev Valery; Kagalovsky Tal;	2011

Table A1. *Cont.*

Title	Journal	Authors	Year
The gains and the pitfalls of reification - the case of algebra.	Educational Studies in Mathematics	Sfard A; Linchevski L;	1994
What Conceptions Have US Grade 4–6 Students' Generalized for Formal and Informal Common Representations of Unknown Addends?	International Journal for Mathematics Teaching and Learning	Switzer J Matt;	2016
Long-term effects of sense making and anxiety in algebra.		Tall David;	2017
Two major difficulties for secondary school algebra students constructing mathematical thinking.		Thomas M O. J;	1995
On mathematics students' understanding of the equation concept.	Far East Journal of Mathematical Education	Tossavainen Timo; Attorps Iiris; Väisänen Pertti;	2011
On developing a rich conception of variable.		Trigueros María; Jacobs Sally;	2008
Reinvention of early algebra.		van Amerom; Barbara;	2004
The Equal Sign: Teachers' Knowledge and Students' Misconceptions	African Journal of Research in Mathematics, Science and Technology Education	Vermeulen Cornelis; Meyer Bronwin;	2017
Misuse of the equals sign: an entrenched practice from early primary years to tertiary mathematics.	Australian Senior Mathematics Journal	Vincent Jill; Bardini Caroline; Pierce Robyn; Pearn Catherine;	2015
Teachers' knowledge of pupils' errors in algebra.		Wanjala E K; Orton A;	1996
Students' understanding of algebraic notation: a semiotic systems perspective.	The Journal of Mathematical Behavior	Weinberg Aaron; Dresen Joshua; Slater Thomas;	2016

References

1. Engelbrecht, J.; Llinares, S.; Borba, M.C. Transformation of the Mathematics Classroom with the Internet. *Zdm* **2020**, 1–17. [CrossRef]
2. Chen, C.-M. Intelligent Web-Based Learning System with Personalized Learning Path Guidance. *Comput. Educ.* **2008**, *51*, 787–814. [CrossRef]
3. Jankvist, U.T.; Niss, M. A Framework for Designing a Research-Based "Maths Counsellor" Teacher Programme. *Educ. Stud. Math.* **2015**, *90*, 259–284. [CrossRef]
4. Bokhove, C.; Drijvers, P. Effects of Feedback in an Online Algebra Intervention. *Technol. Knowl. Learn.* **2012**, *17*, 43–59. [CrossRef]
5. Boell, S.K.; Cecez-Kecmanovic, D. A Hermeneutic Approach for Conducting Literature Reviews and Literature Searches. *CAIS, 34,12* **2014**, *34*. [CrossRef]
6. Rhine, S.; Harrington, R.; Starr, C. *How Students Think When Doing Algebra*; Information Age Publishing, Incorporated: Charlotte, NC, USA, 2018; ISBN 9781641134132.
7. Booth, J.L.; McGinn, K.M.; Barbieri, C.; Young, L.K. Misconceptions and learning algebra. In *And the Rest is Just Algebra*; Stewart, S., Ed.; Springer: Cham, Switzerland, 2017; pp. 63–78.
8. Jupri, A.; Drijvers, P.; van den Heuvel-Panhuizen, M. Difficulties in Initial Algebra Learning in Indonesia. *Math. Educ. Res. J.* **2014**, *26*, 683–710. [CrossRef]
9. Kieran, C. The Learning and Teaching of School Algebra. In *Handbook of Research on Mathematics Teaching and Learning*; Grouws, D., Ed.; Macmillan Publishing Company: New York, NY, USA, 1992; pp. 390–419.
10. Kieran, C. Learning and teaching of algebra at the middle school through college levels: Building meaning for symbols and their manipulation. In *Second Handbook of Research on Mathematics Teaching and Learning*; Lester, F.K., Ed.; NCTM-Information Age Publishing: Charlotte, NC, USA, 2007; Volume 2, pp. 707–762.
11. Booth, L. *Algebra: Children's Strategies and Errors. A Report of the Strategies and Errors in Secondary Mathematics Project*; NFER-NELSON Publishing company Ltd.: Windsor, UK, 1984.
12. Filloy, E.; Rojano, T. Solving Equations: The Transition from Arithmetic to Algebra. *Learn. Math.* **1989**, *9*, 19–25.
13. Herscovics, N.; Linchevski, L. A Cognitive Gap between Arithmetic and Algebra. *Educ. Stud. Math.* **1994**, *27*, 59–78. [CrossRef]
14. Vlassis, J. The Balance Model: Hindrance or Support for the Solving of Linear Equations with One Unknown. *Educ. Stud. Math.* **2002**, *49*, 341–359. [CrossRef]

15. Booth, J.L.; Koedinger, K. Key Misconceptions in Algebraic Problem Solving. In Proceedings of the Annual Meeting of the Cognitive Science Society, Washington, DC, USA, 23–26 July 2008.
16. Gallardo, A. The Extension of the Natural-Number Domain to the Integers in the Transition from Arithmetic to Algebra. *Educ. Stud. Math.* **2002**, *49*, 171–192. [CrossRef]
17. Herscovics, N.; Linchevski, L. Pre-Algebraic Thinking: Range of Equations and Informal Solution Processes Used by Seventh Graders Prior to Any Instruction. In Proceedings of the PME Conference; The Program Committee of the 18th PME Conference, Assisi, Italy, 29 June–4 July 1991; Volume 2, pp. 173–180.
18. Vlassis, J. Making Sense of the Minus Sign or Becoming Flexible in "Negativity. " *Learn. Instr.* **2004**, *14*, 469–484. [CrossRef]
19. Kieran, C. The Equation-Solving Errors of Novice and Intermidiate Algebra Students. In Proceedings of the Annual Conference of the International Group for the Psychology of Mathematics Education, Noordwijkerhout, The Netherlands, 22–29 July 1985; Streefland, L., Ed.; pp. 141–146.
20. Christou, K.P.; Vosniadou, S.; Vamvakoussi, X. Students' Interpretations of Literal Symbols in Algebra. In *Reframing the Conceptual Change Approach in Learning and Instruction*; Elsevier: Amsterdam, The Netherlands, 2007; pp. 283–297.
21. Brousseau, G. Chapter 5 on didactical contract. In *Theory of Didactical Situations in Mathematics*; Nicolas, B., Martin, C., Rosamund, S.V.W., Eds.; Kluwer Academic Publishers: Dordrecht, The Netherlands, 1997; pp. 246–270.
22. Jankvist, U.T.; Niss, M. The Notion and Role of" Detection Tests" in the Danish Upper Secondary" Maths Counsellor" Programme. In Proceedings of the Tenth Congress of the European Society for Reseach in Mathematics Education, Dublin, Ireland, 1–5 February 2017; Dooley, T., Gueudet, G., Eds.; Institute of Education, Dublin City University and ERME: Dublin, Ireland, 2017; pp. 3825–3832.
23. Falkner, K.; Levi, L.; Carpenter, T. Early Childhood Corner: Children's Understanding of Equality: A Foundation for Algebra. *Teach. Child. Math.* **1999**, *6*, 232–236. [CrossRef]
24. Knuth, E.J.; Stephens, A.C.; McNeil, N.M.; Alibali, M.W. Does Understanding the Equal Sign Matter? Evidence from Solving Equations. *J. Res. Math. Educ.* **2006**, *37*, 297–312.
25. Herscovics, N.; Kieran, C. Constructing Meaning for the Concept of Equation. *NCTM* **1980**, *78*, 9.
26. Rittle-Johnson, B.; Matthews, P.; Taylor, R.; McEldoon, K. Assessing Knowledge of Mathematical Equivalence: A Construct-Modeling Approach. *J. Educ. Psychol.* **2011**, *103*, 85–104. [CrossRef]
27. Matthews, P.; Rittle-Johnson, B.; McEldoon, K.; Taylor, R. Measure for Measure: What Combining Diverse Measures Reveals about Children's Understanding of the Equal Sign as an Indicator of Mathematical Equality. *J. Res. Math. Educ.* **2012**, *43*, 316. [CrossRef]
28. Kieran, C. Concepts Associated with the Equality Symbol. *Educ. Stud. Math.* **1981**, *12*, 317–326. [CrossRef]
29. Alibali, M.W.; Knuth, E.J.; Hattikudur, S.; McNeil, N.M.; Stephens, A.C. A Longitudinal Examination of Middle School Students' Understanding of the Equal Sign and Equivalent Equations. *Math. Think. Learn.* **2007**, *9*, 221–247. [CrossRef]
30. Stacey, K.; MacGregor, M. Learning the Algebraic Method of Solving Problems. *J. Math. Behav.* **1999**, *18*, 149–167. [CrossRef]
31. Behr, M.; Erlwanger, S.; Nichols, E. How Children View Equality Sentences. *PMDC Tech. Rep.* **1975**, *3*, ED144802.
32. Collis, K.F. Operational Thinking in Elementary Mathematics. In *Cognitive Development: Research based on a Neo-Piagetian Approach*; John Wiley & Sons: Hoboken, NJ, USA, 1978; pp. 221–248.
33. Hodgen, J.; Brown, M.; Coe, R.; Küchemann, D. Surveying Lower Secondary Students' Understandings of Algebra and Multiplicative Reasoning: To What Extent Do Particular Errors and Incorrect Strategies Indicate More Sophisticated Understandings? In Proceedings of the 12th International Congress on Mathematical Education (ICME-12), Seoul, Korea, 8–15 July 2012; Sung, J.C., Ed.; International Mathematics Union: Seoul, Korea, 2012; pp. 6572–6580.
34. Prediger, S. How to Develop Mathematics-for-Teaching and for Understanding: The Case of Meanings of the Equal Sign. *J. Math. Teach. Educ.* **2010**, *13*, 73–93. [CrossRef]
35. Linsell, C. A Hierarchy of Strategies for Solving Linear Equations. In *Crossing Divides, Proceedings of the 32nd Annual Conference of the Mathematics Education Research Group of Australasia*; Hunter, R., Bicknell, B., Burgess, T., Eds.; MERGA: Palmerston North, NZ, USA, 2009; Volume 1, pp. 331–338.
36. Carpenter, T.P.; Franke, M.L.; Levi, L. Thinking Mathematically. In *Portsmouth NH Heinemann. Carpentier, A.(2012) Les Approch. les Strat. Gouv. mise en Œuvre des Polit. Éducatives. Educ. Francoph. XL*; Greenwood International: Greenwood, NY, USA, 2003; Volume 1, pp. 12–31.
37. Byrd, C.E.; Mcneil, N.M.; Chesney, D.L.; Matthews, P.G. A Specific Misconception of the Equal Sign Acts as a Barrier to Children's Learning of Early Algebra. *Learn. Individ. Differ.* **2015**, *38*, 61–67. [CrossRef]
38. Küchemann, D. Algebra. In *Children's Understanding of Mathematics: 11–16*; Hart, K., Ed.; Murray: London, UK, 1981; pp. 102–119.
39. Sfard, A.; Linchevski, L. The Gains and Pitfalls of Reification: The Case of Algebra. *Educ. Stud. Math.* **1994**, *26*, 191–228. [CrossRef]
40. McNeil, N.M.; Alibali, M.W. You'll See What You Mean: Students Encode Equations Based on Their Knowledge of Arithmetic. *Cogn. Sci.* **2004**, *28*, 451–466. [CrossRef]
41. Linsell, C. Students' Knowledge and Strategies for Solving Equations. *Find. N. Z. Second. Numer. Proj.* **2009**, *2008*, 29–43.
42. Tahir, S.; Cavanagh, M. *The Multifaceted Variable Approach: Selection of Method in Solving Simple Linear Equations*; Mathematics Education Research Group of Australasia: Freemantle, Australia, 2010.
43. Linsell, C. Solving Equations: Students' Algebraic Thinking. *Find. N. Z. Second. Numer. Proj.* **2008**, *2007*, 39–44.

44. Araya, R.; Calfucura, P.; Jiménez, A.; Aguirre, C.; Palavicino, M.A.; Lacourly, N.; Soto-Andrade, J.; Dartnell, P. The Effect of Analogies on Learning to Solve Algebraic Equations. *Pedagog. An Int. J.* **2010**, *5*, 216–232. [CrossRef]
45. Usiskin, Z. Conceptions of School Algebra and Uses of Variables. *Ideas Algebr. K-12* **1988**, *8*, 19.
46. Hart, K. *Children's Understanding of Mathematics: 11-16*; Murray: London, UK, 1981.
47. Küchemann, D. Children's Understanding of Numerical Variables. *Math. Sch.* **1978**, *7*, 23–26.
48. MacGregor, M.; Stacey, K. Ideas about Symbolism That Students Bring to Algebra. *Math. Teach.* **1997**, *90*, 110–113.
49. Hodgen, J.; Küchemann, D.; Brown, M.; Coe, R. Children's Understandings of Algebra 30 Years on. *Res. Math. Educ.* **2009**, *11*, 193–194. [CrossRef]
50. Burkhardt, H.; Swan, M. Task Design for Systemic Improvement: Principles and Frameworks. In Proceedings of the ICMI Study, Oxford, UK, 22 July 2014.
51. Marton, F. *Necessary Conditions of Learning*; Routledge: New York, NY, USA, 2015; ISBN 1317811941.
52. Watson, A.; Mason, J. Seeing an Exercise as a Single Mathematical Object: Using Variation to Structure Sense-Making. *Math. Think. Learn.* **2006**, *8*, 91–111. [CrossRef]
53. Bokhove, C. Using Crises, Feedback, and Fading for Online Task Design. *PNA* **2014**, *8*, 127–138.
54. Tall, D. Cognitive Conflict and the Learning of Mathematics. In Proceedings of the First Conference of The International Group for the Psychology of Mathematics Education, Utrecht, The Netherlands, 29 August 1977.
55. Vergnaud, G. The Theory of Conceptual Fields. *Hum. Dev.* **2009**, *52*, 83–94. [CrossRef]

Paper C

Elkjær, M., & Hodgen, J. (2022). Operationalising Vergnaud's notion of scheme in task design in online learning environments to support the implementation of formative assessment. *Implementation and Replication Studies in Mathematics Education*, *2*(1). https://doi.org/10.1163/26670127-bja10002

BRILL

IMPLEMENTATION AND REPLICATION STUDIES IN MATHEMATICS EDUCATION 2 (2022) 21–44

IRME

brill.com/irme

Operationalising Vergnaud's Notion of Scheme in Task Design in Online Learning Environments to Support the Implementation of Formative Assessment

Morten Elkjær | ORCID: 0000-0001-6035-882X
Danish School of Education (DPU), Aarhus University, Tuborgvej 164, 2400 Copenhagen, Denmark
me@edu.au.dk

Jeremy Hodgen | ORCID: 0000-0002-9196-4088
Department of Curriculum, Pedagogy and Assessment, IOE, UCL's Faculty of Education and Society, University College London, 20 Bedford Way, London WC1H 0AL, UK
Corresponding author, e-mail: jeremy.hodgen@ucl.ac.uk

Abstract

This paper presents an implementation process model for designing and implementing tasks that provide formative feedback in the online learning environment of mathematics classrooms. Specifically, the model operationalises components of Vergnaud's notion of scheme. The implementation process model features a task sequence guided by controlled variation and a 'dual scheme idea'. Using such a sequence of tasks, this work illustrates how Vergnaud's notion of scheme can be used to aid teachers in hypothesising about their learners' understanding of problems involving linear equations, ultimately providing improved feedback for teachers and improved opportunities for student learning in online environments. In Denmark, the online environment matematikfessor.dk is used by approximately 80% of Danish K-9 students.

The impact sheet to this article can be accessed at 10.6084/m9.figshare.19493846.

Keywords

implementation process model – diagnostic tasks – task design – online learning environment – linear equations

 | DOI:10.1163/26670127-BJA10002

1 Introduction

This paper presents the process of developing a framework for implementing formative feedback for teachers engaging in online learning environments. Online learning environments can play a significant role in the implementation of research-based knowledge in the classroom, as online learning environments are capable of providing feedback directly to end users, teachers and students (Dyssegaard et al., 2017). In addition, these environments can produce substantial assessment data with relatively little effort from teachers. However, little research has investigated how online learning environments can best provide feedback to enable teachers to provide effective feedback to their students. There is an extensive research base on how diagnostic tasks performed and why the tasks are believed to be sensible choices for exploring certain areas of difficulty in learning mathematics (Rhine et al., 2018). However, few papers have systematically addressed the considerations or design principles underlying the construction of diagnostic tasks to provide relevant feedback to teachers.

1.1 *The Promise of Online Learning Environments for Mathematics Education*

Online learning environments allow teachers, students themselves or 'intelligent systems' to tailor the content in the environment to the learning needs of students, and these environments can dynamically assign tasks based on students' previous responses (Steenbergen-Hu & Cooper, 2013). Moreover, these environments can 'mark' students' responses almost immediately, thus providing automatic feedback to both students (Cavalcanti et al., 2021) and teachers regarding students' individual and group performance.

However, the process of providing high-quality feedback is not straightforward. Typically, the user is allowed only certain input types: multiple-choice items or numbers. As a result, most of the tasks in online learning environments are closed. Inferring students' mathematical understanding from such tasks can be difficult because a correct answer may be the result of incorrect, or only partially correct, reasoning. This becomes a significant constraint in task design, particularly when working with algebraic expressions, as it is not possible to prompt students for an algebraic expression without presenting them with a multiple-choice option. One possible solution is for the task designer to create distractors that can be set as viable multiple-choice options. Preferably, such distractors would be chosen based on scientific or experimental findings that qualify the option as a good distractor if it indicates a particular difficulty or well-known misunderstanding. However, even good distractors have limitations since it is usually not possible to probe student thinking to infer a student's reasoning behind a particular response.

As online learning environments become more ubiquitous, it is crucial to develop an explicit understanding of the process of designing tasks for implementation in online environments to enable and support teachers' application of the 'big idea' in formative assessment to better their teaching practices (Black & Wiliam, 2009).

1.2 *Matematikfessor.dk: The Context for This Paper*

Matematikfessor.dk, the environment discussed in this paper, has been running in Denmark for over 10 years, and approximately 80% of Danish schools are subscribers to their services. In Denmark, there are approximately 700,000 students in primary school and lower secondary school combined. On an average day, Danish students provide answers to around 1.5 million tasks on *matematikfessor.dk*. This means that in a Danish school year, on average, 250 million tasks are provided with answers on *matematikfessor.dk*. Online learning environments, such as *matematikfessor*, therefore have access to a large amount of data and can potentially provide valuable feedback to teachers regarding the difficulties that students encounter in learning mathematics. *Matematikfessor.dk* provides students with feedback and suggestions on how a task could have been completed if a wrong answer was given. However, like many other online learning environments, student responses are limited to multiple-choice or numeric inputs.

2 Formative Assessment and Related Issues

There is a great deal of evidence that formative assessment can have a positive impact on learning (Black & Wiliam, 1998). Indeed, formative assessment is one of the most widely adopted teaching and feedback provision strategies worldwide. However, attempts to promote formative assessment have often resulted in teachers facing substantial difficulties implementing these ideas (Smith & Gorard, 2005). To help address this implementation problem, Black and Wiliam (2009) proposed five key strategies to support teachers in enacting the 'big idea' of formative assessment: 'evidence about student learning is used to adjust instruction to better meet student needs — in other words, that teaching is adaptive to the student's learning needs' (Wiliam & Thompson, 2007: p. 15). The relationships between these strategies and different aspects of formative assessment are illustrated in Figure 1.

	Where is the learner going	Where is the learner right now	How to get there
Teacher	**1(a)**: Clarifying learning intentions and criteria for success	**2**: Engineering effective classroom discussions and other learning tasks that elicit evidence of student understanding	**3**: Providing feedback that moves students forward
Students as peers	**1(b)**: Understanding and sharing learning intentions and criteria for success	**4** Activating students as instructional resources for one another	
Students as self-regulated learners	**1(c)**: Understanding learning intentions and criteria for success	**5** Activating students as the owners of their own learning	

FIGURE 1 Aspects of formative assessment
ADAPTED FROM BLACK & WILIAM, 2009: P. 8

In this paper, we mainly focus on Strategies 2 and 3, which spotlight feedback for teachers with the purpose of improving future learning situations for students. These strategies engineer items that elicit evidence of student understanding *for* teachers and thus support teachers in providing feedback that helps students advance forward.

In a critical review of Black and Wiliam's (1998, 2009) approach to formative assessment, Bennett (2011) identified six issues with the implementation of formative assessment in the classroom, of which four are relevant to our analysis: the *definitional* issue, the *domain dependency* issue, the *measurement* issue and the *professional development* issue.

Bennett's definitional issue refers to two contrasting views of formative assessment as an instrument or a process. The instrumental view, common among test publishers and many online learning environments, posits that formative assessment is a test that produces a score or set of scores that has 'diagnostic value'. In contrast, the process view, the focus of this paper, sees formative assessment as producing insights into student understanding that can 'actually [be] used to adapt the teaching to meet student needs' (Black & Wiliam, 1998: p. 140).

In highlighting the domain dependency issue, Bennett (2011) argued that eliciting evidence and providing feedback requires an understanding of mathematical learning and development that goes far beyond the generic. To do this, teachers need a 'reasonably deep cognitive-domain understanding … includ[ing] the processes, strategies and knowledge important for proficiency in a domain, the habits of mind that characterise the community of practice in that domain, and the features of tasks that engage those elements' (Bennett, 2011: p. 15).

In focusing on the measurement issue, Bennett argued that assessment is not simply a process of observing students' responses and noting errors or difficulties. Rather, it is an inferential process that requires teachers to have substantial knowledge and expertise that enables them to make productive 'formative hypotheses' and then to act on these. Bennett (2011) argued that this may involve engaging with the student to probe why the student gave a particular answer. Additionally, the teacher could provide more tasks that attempt to determine a pattern in the answers consistent with the hypothesis.

> It is worth noting that the generation and testing of hypotheses about student understanding is made stronger to the extent that the teacher has a well developed, cognitive-domain model. Such a model can help direct an iterative cycle, in which the teacher observes behaviour, formulates hypotheses about the causes of incorrect responding, probes further, and revises the initial hypotheses. In addition, if the underlying model is theoretically sound, it can help the teacher discount student responding that may be no more than potentially misleading noise (e.g., slips that have no deep formative meaning).
>
> BENNETT, 2011: p. 17

The final issue that Bennett raised was the professional development needed for teachers to develop the 'substantial knowledge [required] to implement formative assessment effectively in classrooms' (Bennett, 2011, p. 20). In this paper, our focus is the design of formative tasks within online learning environments that operationalise the components of Vergnaud's (2009) notion of the scheme to enable teachers to better interpret and respond to learners' errors while overcoming some of the issues mentioned by Bennett (2011). In a later section, we return to the notion of scheme and the functionality of its components, as well as how these might serve as guidance in designing formulations of tasks that address the mentioned issues.

3 Research Question

This research was conducted to respond to the following question: How can the notion of scheme guide the design of diagnostic tasks for implementation in online learning environments, specifically regarding known difficulties with the concept of linear equations and the equals sign, to enable teachers to better interpret or hypothesise about learners' difficulties?

The following sections propose a framework for establishing principles to answer the research question. The notion of scheme in the research question refers to the work of Gérard Vergnaud, who established the theory of conceptual fields, including the notion of scheme as an important concept (e.g., Vergnaud, 2009). We expand on the role of this theory in a later section.

This framework is specifically intended to aid in designing diagnostic tasks that enable teachers to hypothesise about their learners' difficulties. Of specific interest in this paper are the difficulties related to the properties of the equals sign in situations involving linear equations. We further elaborate on these properties and difficulties related to students' comprehension of the properties of the equals sign in the following section.

4 Task Design Principles

This section first introduces some key principles that go into the task design and implementation process to address the posed research question, and we further elaborate on the notion of scheme and its role in the task design. The overarching design principle idea stems from the work by Ahl and Helenius (2018) who presented a situation where a student was asked to calculate the average speed. The student invoked a scheme seemingly capable of handling average speed to some extent but ended up invoking and working with a scheme that incorrectly interpreted average speed using an alternative (and insufficient) scheme involving a different average, namely arithmetic mean. Although the student approached the task with knowledge and procedures (a scheme) connected to working with speed, the student ended up applying a scheme that handled the inappropriate arithmetic mean. Because the student was not able to arrive at a satisfactory solution to the task, he ended up invoking another scheme due to the word 'average' (in Swedish, as in English, the word 'average' is used both when discussing average speed and arithmetic mean). We believe this observation made by Ahl and Helenius is a vital finding and opens up a discussion about how to teach students about situations such as this one.

Inspired by Ahl and Helenius (2018), we formulated a sequence of tasks based on an initial task that invokes two potential paths to a solution, one *expected* and one *preferred*. Unlike the situation considered by Ahl and Helenius (2018), we aimed to formulate situations (a sequence of tasks) in which the two paths to the solution are both viable and should result in a correct answer. To create this task sequence, we drew on descriptions of hypothetical (learner) responses guided by variation theory (Watson & Mason, 2006; Marton, 2015). Applying variation theory (Marton, 2015) when designing tasks for this context makes it possible to meaningfully design a sequence of tasks that may be seen as a whole (Watson & Mason, 2006).

When designing such a sequence of tasks, Watson and Mason (2006) proposed the *controlled variation* of tasks and presented the following elements as key factors:

- Analys[e] of concepts in the conventional canon that one hopes learners will encounter.
- Identif[y] of regularities in conventional examples of [...] concept[s] [...] that might help learners (re)construct generalities associated with the concept. [...]
- Identif[y] of variation(s) that would exemplify these generalities;
- Decide which dimensions to vary and how to vary them;
- Construct exercises that offer micro-modelling opportunities, by presenting controlled variation, so that learners might observe regularities and differences, develop expectations, make comparisons, have surprises, test, adapt and confirm their conjectures within the exercise;
- Provide sequences of micro-modelling opportunities, based on sequences of hypothetical responses to variation, that nurture shifts between focusing on changes, relationships, properties, and relationships between properties. (Watson & Mason, 2006: pp. 108–109)

With this task sequence, we created a series of situations through which a learner might experience the schematic shift from the expected path to the preferred path to the solution. The goal of variation theory in our context is to construct the sequence of tasks so that the expected path to the solution increases in difficulty while the preferred path to the solution remains equally efficient. In our case, the intended shift is an extension of the knowledge of the properties of the equals sign. We established variations in situations through a sequence of tasks where students would get to experience the substitution property of equivalence — that if $a = b$, then a can be substituted for b, and vice versa, in any, to the situation relevant equation. We set up the task sequence so that the expected path to the solution would revolve around a scheme guided by equation-solving strategies. However, the intention was that the students may end up observing the substitution property. Before proceeding with the design principles, we discuss some of the difficulties related to the equals sign.

We created this specific sequence of tasks for the contexts of linear equations and the equals sign and drew on research on learners' errors and difficulties with these concepts (Kieran, 1981; Jones et al., 2012; Rhine et al., 2018). Several programmes have attempted to document specific difficulties in learning mathematics. These projects have contributed to the understanding of many misconceptions and other obstacles related to learning mathematics (Rhine et al., 2018). We adopted the view of Jankvist and Niss (2015) who identified genuine difficulties in learning mathematics as 'those seemingly unsurmountable obstacles and impediments — stumbling blocks — which some students encounter in their attempt to learn the subject' (Jankvist & Niss, 2015: p. 260). One kind of stumbling block that many students experience at the beginning of lower secondary school or when attempting to learn to solve more 'abstract' (Vlassis, 2002) linear equations is the role and interpretation of the equals sign (Kieran, 1981). Jones et al. (2012) argued that to achieve a better understanding of the role of the equals sign, one must learn to substitute one representation for a different representation that is equal to the original one. Many strategies for solving equations exist and are all sensibly tied to different tasks and/or situations. However, at some point, even linear equations can become abstract or complicated to the extent where only one strategy is truly viable. Many of these strategies, such as 'guess and check' or 'working backwards', do not necessarily require a deep understanding of the role or interpretation of the equals sign (Linsell, 2009). However, to apply more advanced equation-solving strategies, students must become more flexible in their understanding of the equals sign (Matthews et al., 2012; Kieran, 1981; Rhine et al., 2018).

We utilised controlled variation in the construction of each task in the sequence. By doing this, we attempted to address the importance of learning about the properties of the equals sign mentioned in the previous section. This made it possible to develop a meaningful sequence of tasks that could be evaluated as a whole. When discussing this sequence of nearly similar tasks, Bokhove (2014) suggested that *the element of crisis* is an important factor. Such a crisis occurs when the student completing a range of tasks — or parts of the task range — encounters a task that is impossible or nearly impossible to solve. This element of crisis resembles a cognitive conflict (e.g., Tall, 1977) or an inadequate conceptual field (Vergnaud, 2009). Bokhove and Drijvers (2012) used the element of crisis in their variations when designing a sequence of nearly similar tasks to show that students attempting to solve a crisis-provoking task using strategies appropriate for pre-crisis tasks may result in incompleteness because the earlier strategy is inadequate. In a later section, we return to what we refer to as the 'dual scheme idea', which works implicitly with the element of crisis (Bokhove, 2014).

5 The Notion of Scheme and Its Role in Activity

Vergnaud's (2009) work demonstrates how the scheme as a concept works as an organiser of action or activity when faced with a situation or a class of situations:

> [Schemes] describe ordinary ways of doing, for situations already mastered, and give hints on how to tackle new situations. Schemes are adaptable resources: they assimilate new situations by accommodating to them. Therefore, the definition of schemes must contain ready-made rules, tricks and procedures that have been shaped by already mastered situations.
>
> VERGNAUD, 2009: p. 88

Such a situation or class of situations could be equated to working with algebraic expressions or engaging in solving linear equations. If we accept that schemes are organisers of an individual's activity, we can create assumptions about students' schemes by observing their actions in desired situations. Ahl and Helenius (2018) claimed that this is why schemes are both didactically and analytically more interesting than the idea of conceptual understanding.

Vergnaud (2009) defined a scheme as having four aspects:

> The *intentional aspect* involves a goal or several goals that can be developed in subgoals and anticipations. The *generative aspect* involves rules to generate activity, namely the sequences of actions, information gathering, and controls. The *epistemic aspect* involves operational invariants, namely concepts-in-action and theorems-in-action. Their main function is to pick up and select the relevant information and infer from it goals and rules. The *computational aspect* involves possibilities of inference. They are essential to understand that thinking is made up of an intense activity of computation, even in apparently simple situations; even more in new situations. We need to generate goals, subgoals and rules, also properties and relationships that are not observable.
>
> The main points I needed to stress in this definition are the generative property of schemes, and the fact that they contain conceptual components, without which they would be unable to adapt activity to the variety of cases a subject usually meets.
>
> VERGNAUD, 2009: p. 88, our emphasis on the aspects

Essential to schemes from Vergnaud's perspective are the operational invariants (the epistemic aspect of schemes) consisting of *concepts-in-action* and *theorems-in-action*. A concept-in-action is 'an object, a predicate, or a category that is held to be relevant' (Vergnaud, 1988: p. 168). In every mathematical action, we choose certain objects, predicates or categories that are believed to be relevant to the current situation or setting. A theorem-in-action is a proposition held to be true. When we engage in a mathematical situation, we believe certain 'theorems' to be true or false regarding the objects relevant to the situation. Vergnaud stated that there is a dialectic connection between theorems and concepts, and this emerges from the fact that more advanced mathematical concepts originate from theorems, and vice versa. Nonetheless, it is important to distinguish the cognitive functions of the operational invariants in this precise manner. Concepts-in-action are individually available concepts in a, for the enactor, relevant representation to the situation. We emphasise that Vergnaud's (1988) interpretation of representation is similar to what others call a conception, a concept image or an invoked concept image (Tall & Vinner, 1981). Concepts-in-action bear no value in terms of logical truth, just relevance to the situation. Theorems-in-action are by nature true or false. These entities are sentences (or propositions) that provide the concepts with the possibility of inferences taking place. The *rules of action* are not to be confused with theorems-in-action. The function of the rules of action (the generative aspect of the scheme) is to be appropriate and efficient, but they rely implicitly on theorems-in-action (Vergnaud, 1997). Vergnaud (2009) emphasised that schemes are efficient organisers of activity by nature, and should they also become effective, the scheme can be considered an algorithm. He further clarified that schemes do not have all the characteristics of algorithms. The effectiveness of algorithms allows them to find a solution to a task using a finite number of steps (if a solution is possible).

6 Implementation of the 'Dual Scheme Idea' in a Sequence of Tasks

This section presents the task design and the reasoning behind it. The sequence of tasks was generated using the principles of controlled variation (Watson & Mason, 2006) based on the 'dual scheme idea' with expected and preferred paths to a solution. We also present formulations of a task from the sequence guided by the four components of the scheme (Vergnaud, 2009). An added discussion of the considerations regarding implementation in an online environment precedes each task formulation.

As mentioned, we set up a sequence of tasks (situations), of identical form, where two different schemes might be in play at the same time. The first line in each task can be interpreted as an equation that needs to be solved for the unknown value x. The second line in each task presents an expression for evaluation based on the knowledge acquired in the first line. However, the solution to a task from the sequence does not indicate what path (scheme) the students might use to obtain the solution. The alternate formulations guided by the components of the scheme should reflect which scheme got the upper hand, providing teachers with feedback that can help them generate 'formative hypotheses' regarding their students' schemes. In some cases, we proposed several formulations of the task in an attempt to address different aspects of the components of the scheme. We argue that this 'dual scheme idea', together with the principle of controlled variation, can help task designers better focus the intention and purpose of tasks in online learning environments. These principles lead to a clearer distinction between the expected and preferred paths to a solution and thereby a more explicit articulation of the learning objective. We focused the variations on the value of the unknown. This choice stemmed from the desire to create a sequence of increasingly more 'difficult' tasks based on the expected path to the solution where the preferred path remains at the original level of difficulty. We remind the reader of the role of the epistemic aspect of the scheme. Focused on the operational invariants, this aspect is such an essential part of the scheme that it and its elements (concepts-in-action and theorems-in-action) inform or are present in the other three components. For example, one would simply not be able to establish goals without having at least partial access to a concept relevant to the situation — in other words, a concept-in-action. The focus of the formulations guided by the components of the scheme is on evaluating whether students invoke schemes capable of handling substitutions based on equality, or rather, a scheme suitable for solving equations to solve the task and thus provide teachers with information on students' schemes.

1. What number should go into the empty space?
   $3x = 9$
   $3x + 4 =$ ___
2. What number should go into the empty space?
   $4x = 10$
   $4x + 3 =$ ___
3. What number should go into the empty space?
   $3x = 2$
   $3x + 4 =$ ___
4. What number should go into the empty space?
   $3x = 11$
   $3x + 4 =$ ___
5. What number should go into the empty space?
   $3x = -6$
   $3x + 4 =$ ___

Applying a scheme for handling and substituting equal terms will, in most cases, be the 'easier' path to the goal of solving the above tasks. The term $3x$ can be treated as an object. An expert equation solver would choose the most efficient scheme to solve the task: 'I know x has a specific value and I could easily calculate it, but I do not need to for this task because the object $3x$ is present in both equations'. However, many students believe that such tasks need to be solved using the method that they have been taught, and then they encounter difficulties implementing it. The sequence of tasks is designed to target and isolate these difficulties. The expected path to the solution among students involves solving the first equations for x and then substituting the value of x into the second equation, allowing them to do calculations to fill in the empty space.

In Task 1, a lower secondary school student might look at it and just 'know' the value of x (Linsell, 2009) or simply ignore the letter (Küchemann, 1981). In our experience, many students in lower secondary school become able to read an equation, such as the top equation in Task 1, and imagine it as a command that sorts out the value of a spot in a multiplication table. Traditionally, students around the world would be familiar with tasks such as $4 + 3 =$ ___, where the goal is simply to fill in the empty space, and they would do so by adding the numbers on the left side of the equals sign. Similarly, students would have experienced tasks such as $3 \cdot$___ $= 12$, having worked with all four basic operations. The difference in this task sequence, and a difference one should be aware of, is the missing multiplication symbol between the coefficient and the unknown. From previous work, we know that expressions such as $3x$ are easily misinterpreted by students (Rhine et al., 2018).

In Task 2, a lower secondary school student might still be able to guess/know the value of x. However, if the student is not comfortable or familiar with the idea that non-integer numbers could be solutions to an equation, such as the ones presented in this context, the student would have to alter their strategy based on this crisis or cognitive conflict. The coefficient to x is not a true divisor of the number to the right of the equals sign. This does not necessarily mean that a student in lower secondary school does not know or is able to guess the value of the unknown.

In Task 3, we go a step further. To determine the value of x, the student should be comfortable with numbers between 0 and 1. If the student is also uncomfortable with fractions, they would end up with an infinite decimal number. We expect that most students will experience difficulties knowing or guessing the value of the unknown at this stage.

Task 4 is in many ways like Task 3. However, in this task, the students solving the equations deal with an improper fraction or an infinite decimal number.

Task 5 changes the situation a little bit. As a final task in the sequence, we let the unknown be a negative number. Vlassis (2002) demonstrated that linear equations become significantly more difficult when they involve negatives. In this task, we chose *not* to have the coefficient take a negative value because this might make the task more difficult than having the multiplication result in a negative number and because we wanted to simply vary the value of the unknown.

The common idea among all these tasks is that if students can take the preferred path by substituting the right side of the top expression for the term with the unknown in the bottom expression (in the first tasks, substitute $3x$ for 9), our assumption is that most secondary school students can then straightforwardly solve the task sequence. As mentioned, the tasks were designed so that a 'correct' answer might obscure some difficulties related to linear equations and substitution and therefore are items that generate answers that may be misinterpreted by teachers. If students are more inclined to determine the value of x, because that is what is 'expected' when faced with linear equations, then they might face, as the difficulty of solving the equations increases through the sequence due to the way the numbers were chosen. Each task becomes gradually more difficult if a student attempts to calculate the value of the unknown number x. We realised that the students could reach a fraction as the solution in every task. To solve the task (finding the number that should go in the empty space), students would first have to do the division and end up with increasingly more complex fractions. The same number that just served as the divisor should then be multiplied onto these fractions, resulting in the same number that served as a dividend. However, if students pursue this path and do not preserve the fraction, multiplication could lead to tricky situations dealing with infinite decimal numbers. In addition, a student able to solve the sequence of equations for x (accepting fractions as solutions) should eventually realise that, when inserting the value of x in the second line, they will arrive at the same number that was just on the right side of the first equation and begin to make inferences about substitutional properties.

In *matematikfessor.dk*, tasks must meet certain criteria regarding structure and user input types. The input types are restricted to inputting either a number or multiple-choice selection. Each task must have a unique 'right answer' and must be presented in such a way as to make immediate feedback possible. The tasks presented in this paper are suited for formative and educational purposes rather than training exercises, a categorisation that pertains to many other short-in-formulation equation-solving tasks.

Before we present the formulations guided by the components of the scheme, we emphasise our methodological intention for future teachers working with the task sequence and the formulations based on it. With the overall aim being the implementation of new possibilities for formative feedback for teachers using online learning environments, the intention of this framework is not to leave the teachers without any instructions associated with the task sequence. When working with a task sequence designed using hypothetical responses to variations that nurture shifts between focuses, leading to potential new learning (adaptations in schemes), teachers are provided with a clear common learning goal. Therefore, working with task sequences, such as the above, can lead to possibilities for classroom discussion of the differences in paths to solutions the students might have taken. Alternatively, the task formulations guided by the components can help teachers hypothesise about learners' schemes and formulate a basis for classroom discussion. In our opinion, the latter provides teachers with an opportunity to work with schemes and their components as a more practical tool for teaching problem-solving or engaging in mathematical situations in general.

6.1 *Using the Components of the Scheme as a Guide for Asking Formative Questions*

With the sequence of tasks established, we now present alternate formulations guided by the components of the scheme. In the following sections, we go through the four components (intentional, generative, epistemic and computational) to make alternative formulations of the tasks from the sequence that enable teachers to hypothesise about their students' schemes and difficulties.

6.2 *Setting 'Goals and Anticipations' (the Intentional Part of the Scheme)*

Under this category, we present formulations where setting a goal for or anticipation of the task forms the solution. Before choosing a strategy to arrive at a solution, one must set a goal for what solves the task and what is expected to arrive at a solution. In this intentional aspect of the scheme, one also establishes what the task (the situation) anticipates and what one anticipates from the situation. In many cases, one could expect that, when confronted with a task containing a linear equation, finding the unknown value would be crucial. We propose that when focusing on the goals and anticipation part of the scheme, the task could be formulated as follows (we demonstrate this by using the values from initial Task 4):

Is it necessary to know the value of x to fill in the empty space?

$3x = 11$

$3x + 4 = \_\_\_$

Alternative wording could be as follows:

> Is it *beneficial* to know the value of x to fill in the empty space?
> Would the task be much easier if you knew the value of the unknown x?

Goals and rules are set and established based on the concepts-in-action and the theorems-in-action (the epistemic aspect of the scheme). Whether a student takes note of the equality between $3x$ and 11 and the fact that the term $3x$ is present in both equations as relevant information (concepts-in-action) could indicate whether they are capable of substituting the term $3x$ for 11 in the two equations. Should students not choose the substitutional link between the two equations to be relevant, we expect that they would argue that they would like to know the value of x to fill in the empty space. One might argue that the goal of the task is blurred by the new formulations, since there is an empty line 'begging' for a number to be put on it, but the task is answered by a simple yes or no. However, for the purpose of providing feedback to teachers, following our intention to focus on learning according to Strategies 2 and 3 (Black & Wiliam, 2009), teachers might learn about their students' schemes with this formulation as opposed to just receiving a correct or incorrect answer from students filling in the empty space.

The students might expect that if the unknown value were provided, the value would 'make sense' to the situation and therefore be a small natural number because the numbers present are such. Even if a student were to choose to answer yes to one of the formulations, the student would most likely not expect to be provided an improper fraction as the value of the unknown.

6.3 *Applying 'Rules of Action' (the Generative Part of the Scheme)*

Working with this component of the scheme, we attempted to uncover the rule, or strategy, that students would apply to solve the tasks considering the two schemes. As mentioned in the above category, operational invariants help set the goal or apply rules. After the student has established a goal and/or anticipations, the student can choose appropriate rules to generate action. When forming a rule or a strategy to solve a task, theorems-in-action might be more in focus. Therefore, we formulated the task not in terms of concept relevance but rather in theorems-in-action that lead to rules of action. To confirm that the first equation presented in each of the initial tasks (e.g., $4x = 10$) is in fact important in filling in the empty space, we propose a formulation of the tasks that hints at what 'path to the goal' a student would rather choose: an equation-solving strategy or a substitution of equal terms strategy.

How is $3x = 11$ important in filling in the empty space?

$3x \quad = 11$

$3x + 4 = \_\_\_$

1) Because $3x$ and 11 can be substituted in both lines.
2) Because it lets me calculate the value of x.
3) Another reason.
4) It is not important.

In this formulation, the teacher will get a slightly different view of what path a student wants to choose. A further formulation of the task focused on rules to generate action could appear as follows:

How would you attempt to find the number that goes in the empty space?

1) I would find the value of x.
2) I would substitute $3x$ for 11.
3) I would do something else.
4) I don't know.

If the scheme(s) upon which the student is drawing are only able to apply the rule of action to determine the value of the unknown, the teacher can be provided with valuable information. In this way, a teacher gets a different perspective on essentially the same task but in a different formulation and with a different focus or aim. The rules of action might differ when students engage with the formulations of the different tasks from the sequence due to the increased difficulty when applying an equations-solving scheme.

6.4 *Handling Information with 'Operational Invariants'* (*the Epistemic Aspect of the Scheme*)

With this category, we enter the more complex part of the scheme. Concepts-in-action are concepts relevant to the student engaging in the task, while theorems-in-action are propositions held to be true in a given situation. Therefore, we were careful not to change the situation when we created formulations that attempt to engage with either theorems or concepts related to the task sequence. We remind the reader that in the epistemic part of the scheme, these operational invariants are in play as an underlying aspect of the other parts of the scheme. However, we find it fit to ask questions regarding the relevance of objects or questions regarding the truth value of statements in an attempt to uncover the structure of students' schemes. We begin by addressing the theorem that applies when solving Task 4 to observe whether students might agree to it.

Is it true that the number that goes on the empty line is 4 more than 11?

Another way to address this matter could be via a multiple-choice question:

Can '$3x$' simply be thought of as another way to write '11'?

1) Yes, because that's what the equals sign means.
2) No, because it says that 3 multiplied by some number makes 11.
3) I think both 1 and 2 sounds correct.
4) I do not agree with either 1 or 2.

Or,

Is it okay to substitute $3x$ with 11 in the two lines/equations in the task?

When working with concepts-in-action, we attempted to address the relevance of the objects present in the task. It can be very tricky to ask questions about the relevance of a concept in a given situation. We remind the reader that, in this context, the aim is to assess what students consider relevant objects/concepts.

Is it okay not to care what the value of the unknown is when filling in the empty line?

If the scheme(s) upon which the student is drawing are only able to determine what number that go into the empty space by performing calculations with the value of the unknown number, we expect students not to instinctively agree with such statement.

6.5 *Generating Space for 'Possible Inferences' (the Computational Aspect of the Scheme)*

In this last example, we attempt to address the computational part of the scheme with another formulation of the initial task. Specifically, this formulation aims to determine whether a student makes inferences about the element $3x$ when they compare it to a similar-looking task, but where the scheme for solving equations should be rendered useless because $3x$ has been replaced by a blue box.

Does the same number go into the empty space in both tasks?

$3x = 11$

$3x + 4 =$ ___

☒ = 11
☒ + 4 = ___

If a student does not see the similarities between the two tasks and uses a scheme to solve equations by working with the leftmost task and substituting the blue box and the number 11 in the rightmost task, the student might become suspicious. One might expect the student to wonder why this is the case or why the tasks perform differently, yet so similarly. A different formulation could therefore be as follows:

> Is it surprising that the same number goes into the empty space in both tasks?

The reason for wording the task this way is to establish a cognitive conflict (an element of crisis) in students who are surprised that 15 is the correct number for both empty spaces. This formulation attempts to create a link to a scheme we consider to be like the scheme capable of substituting mathematical equal terms by introducing the task with the blue box.

Another idea is to flip the situation to observe whether the student is willing to infer.

> Is it equally difficult to fill in the empty space in both tasks?

$3x = 11$
$3x + 4 =$ ___
$3k + 4 = 15$
$3k =$ ___

The task on the right-hand side resembles a more ordinary equation-solving situation since the second line presents fewer terms than the first.

7 Discussion

This paper explored the considerations and reasoning behind an implementation process model that can guide diagnostic task design for an online learning environment using an illustrative exemplar set of tasks. This exercise operationalised Vergnaud's idea of the scheme to enable or improve teachers' opportunities to interpret or hypothesise about learners' difficulties within the constraints of the online learning environment.

This section presents a more generalised overview of the proposed framework as an *implementation process model* (Nilsen, 2015). The motivation for establishing such model was the desire to explore the potential of asking formative questions guided by components suitable for online environments. Another part of the motivation stemmed from Bennett's (2011) critical review of formative assessment, which identified six major issues with managing or implementing formative assessment in schools. We set out to address three of the major issues, namely the measurement issue, the domain dependency issue and the definitional issue. In the following, we present our condensed implementation model before going into a discussion of how this model addresses the three mentioned issues with formative assessment.

7.1 *Implementation Process Model*

The first step in our model is to establish the task sequence based on an important property or difficulty in mathematics. In this paper, we looked at the property of the equals sign. The second part of the implementation process model involves transforming the task sequence into questions guided by the components of the scheme. This adds a different layer of variation.

To create a sequence of tasks for use in online learning environments that can improve feedback for teachers and students, we propose the following steps as part of the implementation process model:

1. *Identify regularities* in examples of a concept that might *help learners (re)construct the generalities* associated with that concept. This identification could be guided by research findings on the difficulties students experience in the different subject areas of mathematics education.
2. Establish the *dual scheme idea* with the preferred and expected paths to the solution. This step is extremely important as a guide when designing the variations within the task sequence and within the formulations guided by the components of the scheme.
3. Design a *sequence of tasks* using controlled variation.

The second part of the implementation process model becomes the framework for transforming the task sequence into questions related to the four components of the scheme. This step in the model enables improved feedback to teachers working with online learning environments.

4. Formulate questions that are related to the individual *components of the scheme*, with an emphasis on involving the *dual scheme idea*, as

the questions strengthen the opportunities for the intended learning objective to be successful. Furthermore, these alternate formulations enhances the possibilities for teachers to hypothesise about their learners' schemes.

The implementation process model is illustrated in Figure 2. The steps on the left represent the four steps in the model. The difference in feedback is illustrated on the right.

We realise that the 'dual scheme idea' does not necessarily encapsulate all possible objectives of learning. However, from a teaching perspective, this helps control the situations and enables a clear path for instruction when discussing how the two different schemes handle the tasks, especially in relation to the four major formative assessment issues of measurement, domain dependency, definitional and professional development (Bennett, 2011). We believe that we are now in the business of providing teachers that use online learning environments with useful classroom materials that integrate pedagogical, domain and measurement knowledge regarding formative assessment.

Regarding the definitional issue, Bennett (2011) argued that seeing formative assessment as a process or a test/instrument is an oversimplification. If formative assessment were to be thought of as an instrument, Bennett argued that a carefully developed, research-based instrument is still unlikely to be effective in instruction if the process surrounding its use is flawed. When working with online learning environments that operationalise the components of the scheme as in our model, we get a step closer to creating an instrument with a functioning process supporting it.

In our implementation process model is a designed framework that accommodates for the measurement issue. This issue might also be the most tangible of the three issues discussed here. The task design, considerations and ideas

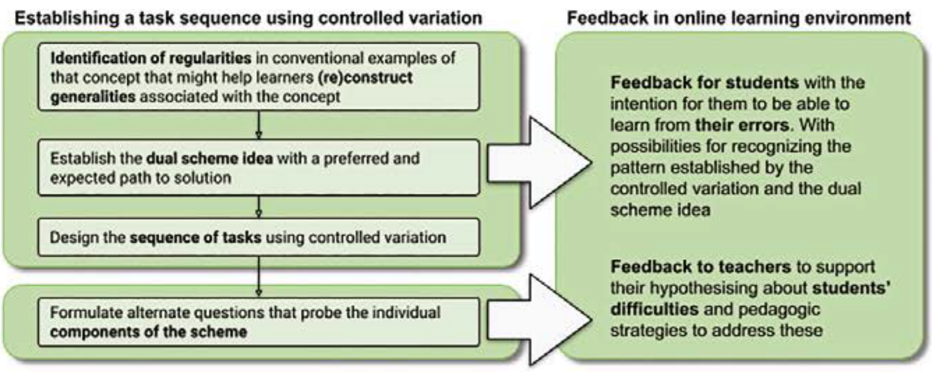


FIGURE 2 Diagram of the implementation process model

behind the implementation process model have been discussed. The next step is to gather feedback from teachers and study teachers' use of this model.

Finally, the use of the components of the scheme enabled us to address the domain dependency issue. Asking questions using the components can provide a deep cognitive-domain understanding that has the potential to reveal general principles, strategies, techniques and knowledge important for proficiency in mathematics.

7.2 *The What, How and Why*

In this paper, we considered just one sequence of tasks to demonstrate how the four components of Vergnaud's idea of the scheme can be operationalised to support teachers in formatively assessing students' understanding within the constraints of an online learning environment. The intentional aspect relates to *what* needs to be done to solve the task or *what* is expected of the student as they attempt to reach the anticipated goal. The generative aspect relates to *how* expectations are met or *how* progress is to be made in the situation. Finally, the computational aspect relates to *why* the desired goal is achieved or *why* new connections to other schemes or concepts make sense or might be established. We have intentionally left out the epistemic aspect of the scheme in this section since the epistemic is such an essential part of the scheme and will be present when working with the other components. We argue that this resembles how clarifying questions asked by teachers should always take form.

The *what, how* and *why* might potentially contribute to the shared or agreed upon theory of change, understood as such that all stakeholders involved with the chain of implementation, so to speak, agree on the framework (Jankvist et al., 2021). This understanding can thereby strengthen the implementability of the tasks, as the idea of respecting the *what, how* and *why* among teachers using this framework with online learning environments will be shared (Jankvist et al., 2021).

8 Concluding Remarks

The next steps for the implementation process model include implementing the model in the design of further tasks in online learning environments, such as *matematikfessor*. It is also important to examine whether, and how, the resulting tasks actually enable teachers to hypothesise about their learners' schemes. We used Vergnaud's components of the scheme to demonstrate how an array of exemplar items can be designed to distinguish between the different schemes that students use to tackle a task.

We believe that Vergnaud's components provide a productive approach to describing students' math-related actions to teachers. We also believe that online environments, such as *matematikfessor*, could help implement this for many teachers. However, our paper highlights an urgent need for work on task design for online environments such as *matematikfessor*. One affordance of working with online environments is that teachers, as well as students, can get easy access to feedback. However, we do not claim that the task of creating a task sequence, such as the one presented in this paper, is necessarily easy. We do, however, believe that this could also be considered an advantage of online learning environments. With professional task designers and a framework for generating tasks with added formative feedback, task sequences might benefit teachers in planning future lessons.

Further steps include the design of additional task sequences in different areas of mathematics education. These sequences of tasks, alongside formulations using the components of the scheme, should then be implemented in an online learning environment, such as *matematikfessor*. Then, a study examining teachers' experiences with the formulations of the sequences using the components of the scheme should take place. Getting feedback on not only the interpretational potential and effect but also the way the teachers go about operationalising the feedback they receive from using the formulations is important. Knowledge sharing and feedback from teachers using the tasks will pave the way for the sensible implementation of a range of sequences of tasks.

Impact Sheet

The impact sheet to this article can be accessed at 10.6084/m9.figshare.19493846.

References

Ahl, L. M. & Helenius, O. (2018). Why is calculating the average speed difficult? In E. Bergqvist, M. Österholm, C. Granberg & L. Sumper (Eds.), *Proceedings of the 42nd Conference of the International Group for the Psychology of Mathematics Education*, Vol. 2 (pp. 3–10). PME.

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

Black, P. & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, *5*(1), 7–74. https://doi.org/10.1080/0969595980050102.

Black, P. & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, *21*(1), 5. https://doi.org/10.1007/s11092-008-9068-5.

Bokhove, C. (2014). Using crises, feedback, and fading for online task design. *PNA*, *8*(4), 127–138. https://doi.org/10.30827/pna.v8i4.6111.

Bokhove, C. & Drijvers, P. (2012). Effects of feedback in an online algebra intervention. *Technology, Knowledge and Learning*, *17*(1–2), 43–59. https://doi.org/10.1007/s10758-012-9191-8.

Cavalcanti, A. P., Barbosa, A., Carvalho, R., Freitas, F., Tsai, Y.-S., Gašević, D. & Mello, R. F. (2021). Automatic feedback in online learning environments: A systematic literature review. *Computers and Education: Artificial Intelligence*, *2*, 100027. https://doi.org/10.1016/j.caeai.2021.100027.

Dyssegaard, C. B., Egelund, N. & Sommersel, H. B. (2017). *A systematic review of what enables or hinders the use of research-based knowledge in primary and lower secondary school.* Danish Clearinghouse for Educational Research; Department of Education; Aarhus University, Aarhus.

Jankvist, U. T., Gregersen, R. M. & Lauridsen, S. D. (2021). Illustrating the need for a 'Theory of Change' in implementation processes. *ZDM — Mathematics Education*, *53*(5), 1047–1057. https://doi.org/10.1007/s11858-021-01238-1.

Jankvist, U. T. & Niss, M. (2015). A framework for designing a research-based "maths counsellor" teacher programme. *Educational Studies in Mathematics*, *90*(3), 259–284. https://doi.org/10.1007/s10649-015-9629-8.

Jones, I., Inglis, M., Gilmore, C. & Dowens, M. (2012). Substitution and sameness: Two components of a relational conception of the equals sign. *Journal of Experimental Child Psychology*, *113*(1), 166–176. https://doi.org/10.1016/j.jecp.2012.05.003.

Kieran, C. (1981). Concepts associated with the equality symbol. *Educational Studies in Mathematics*, *12*(3), 317–326. https://doi.org/10.1007/BF00311062.

Küchemann, D. (1981). Algebra. In K. Hart (Ed.), *Children's understanding of mathematics: 11–16* (pp. 102–119). Murray, London.

Linsell, C. (2009). A hierarchy of strategies for solving linear equations. In R. Hunter, B. Bicknell & T. Burgess (Eds.), *Crossing divides: Proceedings of the 32nd annual conference of the Mathematics Education Research Group of Australasia* (Vol. 1, pp. 331–338). MERGA.

Marton, F. (2015). *Necessary conditions of learning*. Routledge, Abingdon.

Matthews, P., Rittle-Johnson, B., McEldoon, K. & Taylor, R. (2012). Measure for measure: What combining diverse measures reveals about children's understanding of the equal sign as an indicator of mathematical equality. *Journal for Research in Mathematics Education*, *43*(3), 316–350. https://doi.org/10.5951/jresematheduc.43.3.0316.

Nilsen, P. (2015). Making sense of implementation theories, models and frameworks. *Implementation Science, 10*, 53. https://doi.org/10.1186/s13012-015-0242-0.

Rhine, S., Harrington, R. & Starr, C. (2018). *How students think when doing algebra*. Information Age Publishing, Charlotte, NC.

Smith, E. & Gorard, S. (2005). 'They don't give us our marks': The role of formative feedback in student progress. *Assessment in Education: Principles, Policy & Practice, 12*(1), 21–38. https://doi.org/10.1080/0969594042000333896.

Steenbergen-Hu, S. & Cooper, H. (2013). A meta-analysis of the effectiveness of intelligent tutoring systems on K–12 students' mathematical learning. *Journal of Educational Psychology, 105*(4), 970–987. https://doi.org/10.1037/a0032447.

Tall, D. (1977). Cognitive conflict and the learning of mathematics. In *Proceedings of the First Conference of The International Group for the Psychology of Mathematics Education*. PME. https://homepages.warwick.ac.uk/staff/David.Tall/pdfs/dot1977a-cog-confl-pme.pdf.

Tall, D. & Vinner, S. (1981). Concept image and concept definition in mathematics with particular reference to limits and continuity. *Educational Studies in Mathematics, 12*(2), 151–169. https://doi.org/10.1007/BF00305619.

Vergnaud, G. (1988). A comprehensive theory of representation for mathematics education. *The Journal of Mathematical Behavior, 17*(2), 167–181. https://doi.org/10.1016/S0364-0213(99)80057-3.

Vergnaud, G. (1997). The nature of mathematical concepts. In T. Nunes & P. Bryant (Eds.), *Learning and teaching mathematics: An international perspective*. Psychology Press, Hove.

Vergnaud, G. (2009). The theory of conceptual fields. *Human Development, 52*(2), 83–94. https://doi.org/10.1159/000202727.

Vlassis, J. (2002). The balance model: Hindrance or support for the solving of linear equations with one unknown. *Educational Studies in Mathematics, 49*(3), 341–359. https://doi.org/10.1023/A:1020229023965.

Watson, A. & Mason, J. (2006). Seeing an exercise as a single mathematical object: Using variation to structure sense-making. *Mathematical Thinking and Learning, 8*(2), 91–111. https://doi.org/10.1207/s15327833mtl0802_1.

Wiliam, D. & Thompson, M. (2007). Integrating assessment with learning: What will it take to make it work? In C. A. Dwyer (Ed.), *The future of assessment: Shaping teaching and learning*, 1st edn. (pp. 53–82). Routledge, Abingdon.

Paper D

Elkjær, M., & Hodgen, J. (2022). Replicating a study with tasks associated with the equals sign in an online environment. In C. Fernández, S. Llinares, A. Gutiérrez, & N. Planas (Eds.), *Proceedings of the 45th Conference of the International Group for the Psychology of Mathematics Education*: *Vol. 2* (pp. 251–258). PME.

REPLICATING A STUDY WITH TASKS ASSOCIATED WITH THE EQUALS SIGN IN AN ONLINE ENVIRONMENT

Morten Elkjær[1] & Jeremy Hodgen[2]

[1]Danish School of Education (DPU) Aarhus University,

[2]University College London (UCL)

This paper presents a case study of a conceptual replication study. We replicated the famous and widely cited task presented in Falkner et al. (1999), 8+4=__+5. In contrast to the original study, we administered the task with the same age group (Grade 6) in a different system (Denmark) and via a large-scale online learning environment (OLE), with a larger sample and two decades later. Our replication indicates that the Danish students performed very significantly better than the students in the original study. We discuss why this is the case and argue that OLEs such as the one we used provide an important opportunity to replicate, and thus better understand, similar results.

INTRODUCTION

There is an increasing interest in replication studies in mathematics education at PME (e.g. Inglis et al., 2018) and beyond (e.g. Jankvist et al., 2021). This interest stems from the replication crisis in psychology research, which has highlighted a large proportion of false-positive results (e.g., Open Science Collaboration, 2015). In part, this may be due to the high degree of flexibility in quantitative and experimental researchers' analytic and design choices (Simmons et al., 2011). The imperative for replication studies in mathematics education is, however, broader than this. Aguilar's (2020) literature review highlights the majority of studies published even in respected mathematics education journals are small-scale and hence influenced by contextual factors that are poorly understood. Hence, replication can perform a crucial function in deepening and extending the validity of findings, because "[t]hrough variations to studies, we can delineate the bounds of the original study's findings" (Melhuish & Thanheiser, 2018, p. 106). Jankvist et al. (2021) emphasises that replication studies are important in the mathematics education community because they enable a more deeply understanding of the phenomena and results. Replication studies can help clarify under which conditions a particular finding is true or not and replication whether the results are stable over time, across different educational systems or different populations (e.g. Cai et al., 2018). Aguilar (2020) concludes that knowing more about the conditions that make it possible for a research finding to take place, and the boundaries of where it remains true, advances our research field as it allows us to broaden our understanding of the contextual variables under which the research finding occurs. This in turn has direct implications for the implementation of research findings in practice.

2022. In C. Fernández, S. Llinares, A. Gutiérrez, & N. Planas (Eds.), *Proceedings of the 45th Conference of the International Group for the Psychology of Mathematics Education* (Vol. 2, pp. 251-258). PME.

In this paper, we present the data of a conceptual replication study (Aguilar, 2020) of the study presented in Falkner et al. (1999) and Carpenter et al. (2003), reporting findings from the use of their famous task $8 + 4 = \_ + 5$. This result that is widely cited in the literature on equivalence (e.g. Knuth et al., 2006). From the study presented in Falkner et al. (1999) we learn that an entire range of 145 sixth grade students provided 12 and/or 17 as the number that should go in the empty space. Students argue that 12 is the answer, because the numbers on the left together makes 12, neglecting the meaning of the +5 on the right side and reflecting an operational, rather than a relational, understanding of the equal sign (Knuth et al., 2006). Others argue that what goes in the empty space is the value of all the numbers added resulting in 17. In the two original studies, we are presented with the following data;

Percent of children offering various solutions to 8 + 4 = □ + 5

Grade	Answers Given					Number of Children
	7	12	17	12 and 17	Other	
1	0	79	7	0	14	42
1 and 2	6	54	20	0	20	84
2	6	55	10	14	15	174
3	10	60	20	5	5	208
4	7	9	44	30	11	57
5	7	48	45	0	0	42
6	0	84	14	2	0	145

Table 1: Data from answer provided to $8 + 4 = \_\_ + 5$ (Falkner et al., 1999, p. 223)

We have recreated the above task with two additional variations $4 + \_ = 7 + 5$ and $6 + \_ = 4 + 5$, and implemented them in a Danish OLE called *matematikfessor.dk*. The variations are made in order to investigate the bounds of Falkner et al.'s (1999) findings. The first variation uses the same format as the original task but the empty space has been moved to the left side of the equals sign. This is done in order to investigate how willing students are to put the number 3, completing the sum $4 + 3 = 7$, ignoring the number 5 at the end, similarly to the original task. We did however not expect the students to be willing to put in 16 (the total sum of the numbers present) but were curious whether the students would put 12 completing the sum on the right side ($12 = 7 + 5$). The third variation also features the empty space on the left side of the equals sign. In this variation we wanted to investigate what numbers students were willing to put in when the number completing the sum disregarding the last number, should be a negative number. We expected this encourage students to view the equation as more of a whole, thereby including the +5 at the end, because negative numbers might be an unacceptable answer or option (Vlassis, 2002).

The context: Matematikfessor.dk an online learning environment for mathematics

In Denmark, as in many other systems, teachers and students increasingly use OLEs. *Matematikfessor.dk*, the environment discussed in this paper, has been running for over 10 years. More than 500,000 students in primary and lower secondary schools have access to the environment and, on a typical day, 45,000 unique students use the variety of tasks offered by the site, and collectively answer 1,500,000 tasks. OLEs like *matematikfessor.dk* therefore have access to a large amount of data and can quickly host replications of tasks such as the ones presented in the sections above in order to generate large amounts of responses. This leads us to the following research question;

What similarities and differences do we see more than 20 years after the original study when implementing the task presented in Falkner et al. (1999) *in an OLE?*

THEORETICAL BACKGROUND

In this section, we collect research about students' conception of the equals sign and comments on the difficulties that emerge from these conceptions. Rittle-Johnson et al. (2011) gives four levels of interpretations of or four meanings to apply to the equals sign in given situations (see table 2).

Level	Description	Core equation structures
Level 4: Comparative relational	Successfully solve and evaluate equations by comparing the expressions on the two sides of the equal sign, including using compensatory strategies and recognizing that performing the same operations on both sides maintains equivalence. Recognize relational definition of equal sign as the best definition.	Operations on both sides with multidigit numbers or multiple instances of a variable
Level 3: Basic relational	Successfully solve, evaluate, and encode equation structures with operations on both sides of the equal sign. Recognize *and generate* a relational definition of the equal sign.	Operations on both sides, e.g.: $a + b = c + d$ $a + b - c = d + e$
Level 2: Flexible operational	Successfully solve, evaluate, and encode atypical equation structures that remain compatible with an operational view of the equal sign.	Operations on right: $c = a + b$ or No operations: $a = a$
Level 1: Rigid operational	Only successful with equations with an operations-equals-answer structure, including solving, evaluating, and encoding equations with this structure. Define the equal sign operationally.	Operations on left: $a + b = c$ (including when blank is before the equal sign)

Table 2. 'Construct Map for Mathematical Equivalence Knowledge' (Rittle-Johnson et al., 2011, p. 3).

One of the central difficulties that students encounter in the transition from an arithmetic thought process to an algebraic one is that they continue to view the equals sign as a 'do something" signal' (Kieran, 1981), or they maintain an urge to 'calculate', out of habit (Alibali et al., 2007). In the context of the task chosen for this study, children do need to be able to consider the right side of an expression involving an equals sign as an expression in its own right. In the words of Rittle-Johnson et al. (2011) an operational view or meaning attached to the equals sign. The main purpose of the task ($8 + 4 = \_ + 7$) is to determine what interpretation of the equals sign a student would apply. In the earlier years in school mathematics students might perceive the equals sign as indication for that calculations has to be made and that the operations on the left side results in a single number on the right side of the equals sign (Alibali et al., 2007; Kieran, 1981).

METHODOLOGICAL CONSIDERATIONS

In August 2020 we implemented the task from Falkner et al. (1999) in the OLE *matematikfessor.dk* as parts of three sets of formative tasks, with a total of 49 unique items about linear equations. The sets were only available for teachers to assign to their students, not for students to find on their own within the environment. A promotion campaign was established in order to notify the teachers subscribing to the services of *matematikfessor.dk* of the formative sets existence and applications. The data (in the form of unique answers) was extracted from *matematikfessor.dk's* database on the 4th of November 2021.

DATA RESULTS

The original task 8+4=__+5

In total 2345 answers were given to the original task presented in Falkner et al. (1999) when we implemented out version in the OLE. In a review of these, we found that only 92 of these answers were from students solving the task multiple times. In table 3 is an overview of the answers the students provided. (64 total answers were omitted. These answers were belonged to a range of 16 additional groups of answers that were less than 1% of the answer total)

Answer	Freq	%
7	1546	65.9
17	363	15.5
12	343	14.6
3	29	1.2

Table 3: Overview of the answers to the task 8+4=__+5

Answer	Freq	%
7	501	62.9
17	143	17.9
12	119	14.9
3	8	1.0

Table 4: Overview of the answers to the task 8+4=__+5 (age 12 and 13) We examined how 12 to 13 year olds (6th graders) from Denmark answered the task in order to be able to compare with the same age group from the original study. In total 797 students from this age group answered the implementation of the original task. The results can be seen in table 4.

The amount of 12 year olds that gave the answer 7 is 57.3% where the 13 year olds sum up to 64.0%. The average age of the children represented in the data for the original task is 13.97 years, slightly lower than the total average age of 14.08 years of the children represented in all three tasks. See age distribution in figure 1.

The first variation 4+__=7+5

For the second task (the first variation), we received a total of 1203 answers. In a review of these, we found that only 40 of these answers were from students solving the task multiple times. In table 5 is an overview of the answers the students provided. (45 total answers were omitted. These answers were belonged to a range of 14 additional groups of answers that were less than 1% of the answer total)

Answer	Freq	%
8	996	82.7
3	99	8.2
9	35	2.9
12	27	2.2

Table 5: Overview of the answers to the task $4 + \_ = 7 + 5$

The second variation 6+__=4+5

For the third task (the second variation), we received a total of 824 answers. In a review of these, we found that only 43 of these answers were from students solving the task multiple times. In table 6 is an overview of the answers the students provided. (27 total answers were omitted. These answers were belonged to a range of 13 additional groups of answers that were less than 1% of the answer total)

Answer	Freq	%
3	751	94.9
4	14	1.8
-2	11	1.4
9	11	1.4
2	10	1.3

Table 6: Overview of the answers to the task $6 + \_ = 4 + 5$

Additional results

A total of 351 students have provided answers to all three items. Based on these data the facility of the original task is 69.5%. The facility of the first variation is 83.3% and the facility of the second variation is 93.3%. These students actually represent the overall data very well. Only 32 students have provided two answers to one or more of the items where one of the answers were wrong. We thought it might be interesting to know the exact number of students who either got it wrong first and then right and vice versa. Twenty of the students that provided answers to the task $8 + 4 = \_\_ + 7$ provided two answers, where the first answer was wrong and the second answer correct. Most of these cases were a situation where either 12 or 17 was the first answer and 7 the second. Five students did in fact provide a correct answer as the first and a wrong answer the second time around.

DISCUSSION

In this section, we discuss the similarities and differences in data results compared to the original studies. In addition, we discuss what possible influence the OLE have on with the similarities and differences. If we compare the data from the original study presented in Falkner et al. (1999) we immediately notice the striking difference in facility among 6th grade students. In the original study, every 6th grade student gave the wrong answer to the task. A later publication (Carpenter et al., 2003) provides additional information about the performance of the task and interpretations made by the authors. The author's comment that the data show that older students are more likely to get the task wrong than younger students are and the author hint at that maybe students get a progressively more operational interpretation of the equals sign based on the teaching at this point in time. Knuth et al. (2006) emphasises that poor performance on measures of understanding the equals sign should not be surprising given the lack of explicit focus in American middle school curricula, although we note that a recent study indicates that American students may have a better understanding of equivalence more generally than some European countries (Simsek et al., 2021). McNeil (2007) finds that performance on equivalence problems such as the ones discussed in this paper decreases with American students from age 7-9 before it increases again from age 9-11. Hence, performance on this item may be particularly influenced by pedagogic and curricular choices. Nonetheless, the data from our study show that students in 6th grade (12-13 year olds) give a correct answer 63% of the time and matches the overall distribution very well. We acknowledge that the original study does not specifically intend to provide information on how 6th grade students perform on a task such as $8 + 4 = \_\_ + 7$. Rather they intend to provide teachers with a reminder that students' interpretation of the equals sign is of great importance and does not need to be corrected at an older age rather than classroom discussions about the meaning (definition) of the equals sign at lower grades are particular meaningful (Carpenter et al., 2003).

We do get the same wrong answers in our study as in the original. This to some extent prove that the task is not performing in a significantly different way i.e. producing different answers than 20 years ago. We do however wonder why we see the huge difference in the distribution of the answers. 20 years ago in the original study, less than 10% of the participants at every class level gave the answer 7. Now we see a rate of approximately 65%. Granted our data stems from 12-17 years old. With most of the participant being 13-15 (83%). Falkner et al. (1999) mentions that the task was originally carried out by a teacher in a single classroom. When this teacher realized that every student in that classroom provided a wrong answer, she asked her colleagues to use the task with their students resulting in the data in table 1. This means that the observations all stem from the same school. In our study, the data stems from at least 197 schools due to the task being implemented in an OLE. We are however not certain that none of the students in our study received help solving the task. This fact might skew the correct answer percentage towards a higher number. However, it seems unlikely that this should leave us with 60+% correct answers compared to none or almost none. Another obvious difference is nationality of the populations observed in the original study we have American students and in our study the observations stem from Danish students. According to PISA 2018 (https://factsmaps.com/pisa-2018-worldwide-ranking-average-score-of-mathematics-science-reading/) the overall difference in the performance of students in the United States and Denmark is not statistically (or indeed practically) significant. Of course, the task was presented to the American grade 6 students more than 20 years ago and it may be that teachers are now more aware of student's understandings of, and misconceptions about, the equals sign, because of the curricular changes made as a result of the introduction of mathematical competencies in Denmark in 2002. The data collected on the variations of the original task suggests that a similar operational view of the equals sign is being applied even though the empty space is moved to the left side of the equation. This was to be expected, as it is still possible to apply the same operational procedure as the original problem with the empty space on the right side. With the last task, we see an even better performance. The last variation is as expected not similar to the second variation because -2 is not as frequent as the number 3 was in the second variation. This to some extent proves that the choice of numbers matter when designing tasks such as the original task even though the empty space is on the left side of the equals sign. This choice of numbers indicate that students might be more likely to apply a relational interpretation of the equals sign to avoid negative numbers or simply because negative numbers are not accepted in a situation such as this.

CONCLUSION

Based on the differences in the data we believe that, although this task from Falkner et al. (1999), in our opinion is a very good task, the data presented by the authors is not representative of how difficult the task is for 6^{th} grade students. Our data show that the majority of the wrong answers was identical to the ones observed in the original study. This does in our opinion encapsulate one side of the importance of replications studies in mathematics education. On the other hand, our data show a huge deviation from the facility scores of the original study. This is also an important finding for the sake of replication studies in mathematics education. Even though the point of the task presented in Falkner et al. (1999) is not primarily to indicate how difficult it is and present quantitative scores, it is nonetheless important to observe that the scores presented in the original study is an extreme case compared to data collected from a large collection of schools in Denmark 20 years later. With all that said using OLEs to replicate studies such as the performance of the famous task from Falkner et al. (1999) can be great and efficient platforms for achieving additional and in some cases updated information and knowledge.

REFERENCES

Aguilar, M. S. (2020). Replication studies in mathematics education: What kind of questions would be productive to explore? *International Journal of Science and Mathematics Education, 18*(1), 37–50.

Alibali, M. W., Knuth, E. J., Hattikudur, S., McNeil, N. M., & Stephens, A. C. (2007). A Longitudinal Examination of Middle School Students' Understanding of the Equal Sign and Equivalent Equations. *Mathematical Thinking and Learning*, *9*(3), 221–247.

Cai, J., Morris, A., Hohensee, C., Hwang, S., Robison, V., & Hiebert, J. (2018). The role of replication studies in educational research. *Journal for Research in Mathematics Education*, *49*(1), 2–8.

Carpenter, T. P., Franke, M. L., & Levi, L. (2003). *Thinking mathematically*. Portsmouth, NH: Heinemann.

Falkner, K., Levi, L., & Carpenter, T. (1999). Early Childhood Corner: Children's Understanding of Equality: A Foundation for Algebra. *Teaching Children Mathematics*, *6*(4), 232–236.

Inglis, M., Schukajlow, S., Dooren, W. V., Hannula, M. S., & Leuven, K. (2018). Replication in mathematics education. *Proceedings of the 42nd Conference of the International Group for the Psychology of Mathematics Education*, *1*, 2.

Jankvist, U. T., Sánchez Aguilar, M., Misfeldt, M., & Koichu, B. (2021). What to Replicate? *Implementation and Replication Studies in Mathematics Education*, *1*(2), 141–153. https://doi.org/10.1163/26670127-01010015

Kieran, C. (1981). Concepts Associated with the Equality Symbol. *Educational Studies in Mathematics*, *12*(3), 317–326. https://doi.org/10.1007/BF00311062

Knuth, E. J., Stephens, A. C., McNeil, N. M., & Alibali, M. W. (2006). Does Understanding the Equal Sign Matter? Evidence from Solving Equations. *Journal for Research in Mathematics Education*, *37*(4), 297–312.

McNeil, N. M. (2007). U-shaped development in math: 7-year-olds outperform 9-year-olds on equivalence problems. *Developmental Psychology*, *43*(3), 687.

Melhuish, K., & Thanheiser, E. (2018). Reframing replication studies as studies of generalizability: A response to critiques of the nature and necessity of replication. *Journal for Research in Mathematics Education*, *49*(1), 104–110.

Rittle-Johnson, B., Matthews, P., Taylor, R., & McEldoon, K. (2011). Assessing knowledge of mathematical equivalence: A construct-modeling approach. *Journal of Educational Psychology*, *103*(1), 85–104.

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, *22*(11), 1359–1366.

Simsek, E., Jones, I., Hunter, J., & Xenidou-Dervou, I. (2021). Mathematical equivalence assessment: Measurement invariance across six countries. *Studies in Educational Evaluation*, *70*, 101046.

Vlassis, J. (2002). The balance model: Hindrance or support for the solving of linear equations with one unknown. *Educational Studies in Mathematics*, *49*(3), 341–359. https://doi.org/10.1023/A:1020229023965

Paper E

Elkjær, M., & Mørup, M. (submitted). Learning from errors: Co-clustering students' answer types in a digital learning environment to verify task design on linear equations.

Learning from Errors: Co-Clustering Students' Answer Types in a Digital Learning Environment to Verify Task Design on Linear Equations

Morten Elkjær[1] and Morten Mørup[2]

[1]Danish School of Education, Aarhus University, Denmark, me@edu.au.dk, OrciD: 0000-0001-6035-882X

[2]Section for Cognitive Systems, DTU Compute, Technical University of Denmark (DTU), Denmark, mmor@dtu.dk, OrciD: 0000-0003-4985-4368

Abstract

This study is concerned with establishing a means to generate better methods to analyse and learn about task design for digital learning environments. Specifically, we utilise data consisting of over 2 million unique answers from a popular Danish digital learning environment, matematikfessor.dk, to solve 892 unique tasks dealing with linear equations. Utilising the Multinomial Infinite Relational Model (MIRM), which can account for extensive didactical coding of the five most popular answers to each of the 892 tasks, we successfully co-clustered students and tasks into groups for further analysis. The results showed that the analysis of these clusters of tasks can provide access to valuable information on the difficulties students in four respective groups face and what kind of specific tasks and knowledge pertinent to what types of tasks actually cause students' problems or difficulties anticipated by the task designer.

Keywords: Task design; stochastic block modelling; linear equations

Introduction

In a recent book "Digital Technologies in Designing Mathematics Education Tasks", Leung and Baccaglini-Frank (2017) called for additional research into knowledge on tasks involving digital tools. The authors emphasised that:

> *More than a decade ago Sierpinska (2003) identified task design as a core research area in mathematics education. She commented that research reports rarely gave sufficient details about tasks for them to be used by someone else in the same way.* (Leung & Baccaglini-Frank, 2017, p. vvi)

In this paper, we investigate how data from a digital learning environment can aid in how we learn from students' answers and types of errors when solving linear equations. Leung and Baccaglini-Frank (2017) found that seemingly minor differences and variations in tasks could have a significant impact on learning. In addition, digital environments can produce substantial amounts of data in the form of students' answers to tasks with relatively little effort from teachers or researchers. However, little research has been devoted to investigating how the analysis of data from digital learning environments can provide knowledge on how a mathematical task design reflects the structure observed in student responses.

Digital learning environments and digital resources nonetheless allow teachers, students or intelligent tutoring systems to tailor the content in the environment to students' learning needs based on collected data. The intelligent tutoring systems in such environments can sometimes dynamically assign tasks to students based on previous responses (Steenbergen-Hu & Cooper, 2013).

Typically, the user is allowed only certain input types, such as multiple-choice items or numbers, when working in digital learning environments. As a result, most of the tasks in digital learning environments are closed. Inferring students' mathematical understanding based on such data can be difficult because a correct answer may be the result of incorrect or only partially correct reasoning. This becomes a significant constraint in task design and assessment, particularly when working with algebraic expressions, because it is not possible to prompt students for an algebraic expression without presenting them with a multiple-choice option. One possible solution is for the task designer to create distractors that can be set as viable multiple-choice options. Preferably, such distractors would be chosen based on scientific or experimental findings that qualify the options as good distractors if there is indication that they are associated with particular difficulty or well-known misunderstanding. However, even good distractors have limitations, since it is usually not possible to probe student thinking to infer students' reasoning behind particular responses.

It is important to realise that data extracted from digital learning environments will most likely have different characteristics than data from more classic approaches involving students responding to a collection of tasks. In digital learning environments, when attempting to analyse a large collection of tasks, we will not benefit from the fact that each student (seen as a participant or an informant) has answered or interacted with every task in the collection. On the contrary, although we might observe many data points in the form of students' answers to tasks, the data set is far from comprehensive or what we refer to as only partially observed. To clarify, the data matrix (student × task) with entries consisting of answers is very far from fully observed. The entries in a student × task matrix could in fact hold slightly different values, such as correct/incorrect, time spent or the actual answer. Several values could also be included in the analysis. In the case of this paper, we have chosen to perform data enhancement on the actual answers the students provided in the form of reasons why the answers were given. In addition, we aim to utilise a branch of machine learning called unsupervised learning to structure the data in this matrix. Unsupervised learning can structure students' answers in terms of groups that exhibit similar behaviours across groups of tasks. Such groups are often referred to as clusters or co-clusters. The term 'co-clustering' refers to the goal of generating a subset of rows that exhibit similar behaviours across a subset of columns, or vice versa, but simultaneously. Thus, the aim is to co-cluster tasks and students to identify homogenous patterns of answer types optimally. As we do not a priori know the number of groups (clusters) for either axis, we will further explore how Bayesian non-parametric modelling allows efficient inference of a suitable number of groups. The matrix we will use for the co-clustering is a student × task matrix where we have performed data enhancement on all the entries, providing reasons for the answers that would otherwise have served as entries. Furthermore, the student × task matrix is partially observed, with only 2.5% of the entries known. The co-clustering method and the structure of the data enhancement are explained in further detail in the Method section. The data used in co-clustering consisted of 2,135,968 unique answers provided for 892 unique tasks by 94,368 students. Danish students in the digital learning environment matematikfessor.dk provided the answers.

Ultimately, we propose a framework for enhancing the possibilities of using data collected through a Danish digital learning environment to learn about and interpret variables in task design using linear equations. The type of feedback for researchers and task designers that we envision can provide enhanced opportunities for knowledge about pupils' comprehension of a concept and their strategies used to solve linear equations, as opposed to feedback purely based on correctness and some difficulty levels.

Research Question

We hypothesise that prominent groups of student profiles that systematically face similar challenges (defined across groups of tasks with linear equations) can be learned from partially observed collective information acquired from a digital learning environment. Specifically, our investigation aims to answer the following research question:

> *How can the method of co-clustering contribute to the analysis of large partially observed data sets collected in a digital learning environment to generate better possibilities for learning about task design and students' difficulties solving linear equations?*

We emphasise that the choice to enhance the data by providing error codes or anticipated reasons for why the wrong answers were provided in an attempt to move a step closer to our overall goal, which is to provide enhanced knowledge for teachers and task designers dealing with digital learning environments.

We hypothesise that the identification of such groups (i.e., co-clusters), both of students and of tasks, can guide task design for digital learning environments by identifying and understanding the challenges students face when solving linear equations. We further hypothesise the following:

a) We can identify systematic student behaviours at the level of groups, ideally reflecting different prominent competence levels.
b) We can identify task groups that present similar perceived challenges among the identified student groups.
c) We can predict students' performances on unobserved tasks based on their identified group profiles, thereby identifying challenging tasks.

Ultimately, learning the structure of detailed student responses may improve knowledge of task performance in digital learning environments. This knowledge can provide task designers and researchers with extensive insight into task design principles or guide design and task development for digital learning environments.

In the following sections, we go over the assumptions and prerequisites that go into our framework. We first describe how we enhanced the data extracted from the digital learning environment beyond correct and incorrect answers. Subsequently, we explain the Multinomial Infinite Relational Model (MIRM) used for the unsupervised co-clustering of the student × task matrix, which can account for multiple answer types when identifying and analysing coherent student and task groups.

Data

Matematikfessor.dk, the digital learning environment discussed in this paper, has been operating in Denmark since 2009, and approximately 70% of Danish schools subscribe to the services it provides. In Denmark, there are approximately 700,000 students in primary school and lower secondary school combined. On an average day, Danish students provide answers to around 1,500,000 tasks on matematikfessor. This means that in a Danish school year, on average, 250,000,000 tasks are provided with answers on matematikfessor.dk. Digital learning environments, such as matematikfessor, therefore have access to a large amount of data and can provide feedback in the form of statistical overviews. This equips teachers with easy access to both live and summative dashboards regarding their students' performance. It is up to the teachers to infer the difficulties that students encounter in learning mathematics while solving problems in the digital environment, since the statistics are purely based on the correctness of the answers.

Elkjær and Jankvist (2021) reviewed the literature on students' difficulties when solving linear equations, leading to a systematic set of design principles for the structure of linear equations to detect these difficulties. Notably, the presented structure builds on the foundation of a standard non-arithmetic equation (Elkjær & Jankvist, 2021; Filloy & Rojano, 1989) (see Figure 1). The system of linear equations is separated into types of linear equations relevant to lower secondary schools. Specific examples of how to design concrete variations were further demonstrated by utilising variations theory (Marton, 2015) to reach a comprehensive set of linear equations of respective types based and on known difficulties that worsen secondary school students' experience when solving linear equations (Elkjær & Jankvist, 2021).

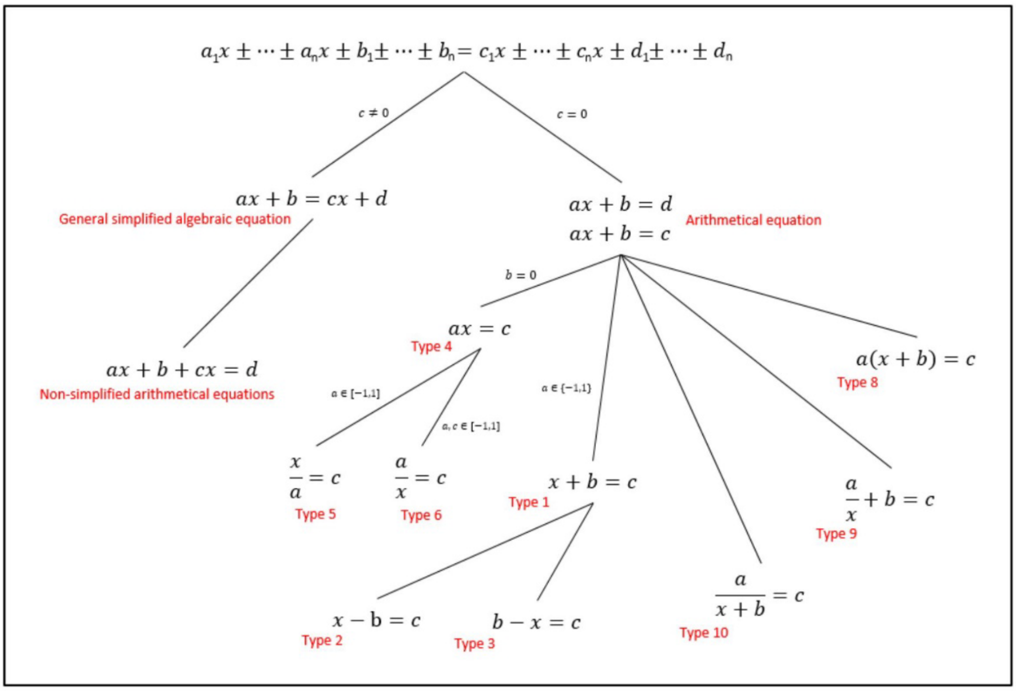


Figure 1: Structure of linear equations adapted from Elkjær and Jankvist (2021, p. 13).

From these difficulties, the overarching design principles were given, leading to the construction of the 892 equations that serve as the basis for data collection. The overall categories of difficulties found in Elkjær and Jankvist (2021) are difficulties related to the following:

- *The concept of numbers*—This involves negative numbers (specifically the role of the minus sign in expressions) and numbers in expressions that belong to sets in general that stretch beyond the natural numbers (i.e., rational numbers and zero) (e.g., Gallardo, 2002; Vlassis, 2002).
- *The equals sign and its role in expressions*—This issue is mainly concerned with how the equals sign is interpreted in concrete equations and how the structure of the expressions is thereby made sense of. In some cases, terms might be disregarded or misread (e.g., Kieran, 1981; Matthews et al., 2012; Prediger, 2010).
- *Strategies and transformations*—Depending on the complexity of the equations, different strategies might come into play, including different sorts of transformations and procedures. This includes conventional concepts, such as the procedures and roles of operators when rearranging or transforming expressions to reach a solution (e.g., Jankvist & Niss, 2015; Kieran, 1985; Linsell, 2009).

- *Letters in expressions*—These issues could be related to the role and handling of coefficients in terms of coefficients being added or multiplied onto the unknown (e.g., Küchemann, 1981).

Incorporating the *element of crisis* (Bokhove, 2017) or what might also be thought of as attempting to ensure that *cognitive conflicts* (Tall & Vinner, 1981) could be addressed while solving equations served as a method for ensuring that the above design principles were incorporated in each of the types of equations. Variation theory guided the choice of numbers present as constants in each type of equation to potentially address as many of the identified difficulties students experience as possible. We demonstrate this concept in Table 1.

Table 1: Variations in equations of Type 4 (adapted from Elkjær & Jankvist, 2021, p. 15).

Original	Variation 1	Variation 2	Variation 3	Variation 4	Variation 5
$3x = 15$ $8x = 8$	$3x = -15$	$-3x = 15$	$-3x = -15$	$15 = 3x$	$28x = 14$ $2x = 9$

The data extracted consisted of 2,135,968 unique answers (of 2,509,352 total answers) provided for a total of 892 tasks by 94,368 students. A total of 373,384 answers were removed as duplicates. We created an additional data set consisting of a subset of the original student × task matrix. The criterion for being included in the subset was that a student should have answered at least 100 tasks. This resulted in the inclusion of 2169 students who had collectively answered all 892 tasks. The reason for choosing to look at a subset was to possibly eliminate some of the weaker signals from students who completed very few tasks.

Going forward with the original data set, whenever we came across a task with two answers from the same student, we removed the oldest entry. The 892 equations were categorised into groups. The groups are presented in Table 2. The equations from the original design that were unfortunately not implemented in the digital learning environment were Type 8 equations.

Table 2: Types of equations and examples

Type of equations	**Example**	**Number of equations**	**Number of answers**
Type 1 (2781)	$3 + x = 8$	95	636.286
Type 2 (2782)	$x - 5 = 8$	54	274.948
Type 3 (2783)	$8 - x = 3$	93	243.649
Type 4 (2784)	$2x = 10$	67	264.301
Type 5 (2785)	$x/2 = 8$	43	78.931
Type 6 (2801)	$18/x = 8$	39	83.838
Type 7 (2787)	$2x - 7 = 11$	90	66.756
Type 7 rev (2786)	$8 = 5x + 2$	99	143.791
Type 9 (2802)	$15/x + 4 = 7$	60	72.514
Type 10 (2803)	$18/(x + 2) = 3$	64	60.563
Non-simplified Arithmetic (2858)	$3x - 7 + 2x = 8$	86	18.470
General Algebraic (2788)	$2x + 5 = x + 10$	102	191.453

Every type of equation from Table 2 came in several variations in an attempt to address as many of the most common difficulties reported in the literature (Elkjær & Jankvist, 2021). Additional variations were part of the collected data. These variations feature tasks of a more diagnostic character, such as $8 + 4 = \_\_ + 5$

(Falkner et al., 1999) and $4 - 8 = \_\_$. Furthermore, these additional tasks featured equations and unknowns in different representations. For example, for one task involving pictures of the balance model with missing cubes on one side, students were asked how many cubes on the other side would be required to achieve balance.

Methods

We curated the data, taking into account the types of errors provided by the students by defining multiple categories beyond correct and wrong answers. We subsequently described the co-clustering procedure used to extract tasks and student groups based on this curated data.

Data Enhancement—Going Beyond Correctness

The overall aim was to generate feedback that is more valuable for teachers using digital resources and digital learning environments. Since the variations in the equations designed as a basis for generating the data stem from reviewing students' difficulties, we could generate the following system of codes for student challenges. This system first pre-analyses the data by gathering the five most popular answers provided for all 892 tasks. These five answers were each coded with one or more tags from Table 3.

Table 3: Codes used in coding the five most popular answers to the included equations

Name	**Tag**	**Explanation**
Correct	c	Correct answer
Negativity issue	n	Answers with additive inverses
One more/one less	o	Answering for Example 6 instead of 5
Multiplication	m	Issues involving multiplication
Addition	a	Issues involving addition
Rearranging for sense-making	r	Rearranging an expression to have it make sense
Equals sign	e	Issues related to the role of the equals sign
Conventions associated to letters	v	Issues related to the interpretations of letters
Solving strategy	s	Issues related to equation-solving strategies
Decimal number	d	Issues related to the acceptance of decimal numbers
Zero	z	Issues related to the number 0
Unknown reason	u	Unable to apply a valuable code

We applied a total of 49 different codes that were combinations of the initial 12 codes. The codes 'correct' and 'unknown reason' were never combined with other codes. In the earlier types of equations, the students were unlikely to make multiple fatal errors visible. However, as the complexity of the expressions grew, the string of errors leading to a particular incorrect answer became longer or more advanced. However, in all five of the most popular answers regardless of the task, the correct answer was always present. Because we were limited to the five most popular answers, the remainder of the answers outside of the top five was coded with 'unknown reason'. The correct answer was always among the top five most popular answers of the 892 equations, but unknown reasons were also present. We were not able to provide one or more of the top five most popular answers with a meaningful reason for incorrect answers 905 times. This is a little over one per task. In fact, 609 of the 892 tasks had unknown reason. In Table 4, we present the distribution of answers that were within the five most popular answers for each task in the categories.

Table 4. Percentage of answers falling under each type of equation

Type of equation	Example	Correct%	Identified%	Unknown%
Type 1 (2781)	$3 + x = 8$	89	5	6
Type 2 (2782)	$x - 5 = 8$	81	12	7
Type 3 (2783)	$8 - x = 3$	79	13	8
Type 4 (2784)	$2x = 10$	84	9	7
Type 5 (2785)	$x/2 = 8$	74	17	9
Type 6 (2801)	$18/x = 8$	89	6	5
Type 7 (2787)	$2x - 7 = 11$	80	10	9
Type 7 rev (2786)	$8 = 5x + 2$	76	12	12
Type 9 (2802)	$15/x + 4 = 7$	73	13	14
Type 10 (2803)	$18/(x + 2) = 3$	89	6	5
Non-simplified Arithmetic (2858)	$3x - 7 + 2x = 8$	67	16	17
General Algebraic (2788)	$2x + 5 = x + 10$	68	14	18

Co-Clustering Using Stochastic Block Modelling

Although many tools for co-clustering exist, (see Govaert & Nadif, 2013 for an overview), we focused on the stochastic block-model (SBM) approach (Holland et al., 1983; Kemp et al., 2006; Mørup et al., 2014; Nowicki & Snijders, 2001; Schmidt & Morup, 2013; Xu et al., 2006). In the context of this study, the SBM was used to optimally co-cluster data according to row and column groups defining homogenous blocks corresponding to systematic consistent answer patterns across groups of tasks within each extracted student group. Specifically, we implemented the Infinite Relational Modelling (IRM) generalisation (Kemp et al., 2006; Xu et al., 2006) that makes use of Bayesian non-parametric modelling as defined through the Dirichlet process, which induces a distribution over partitions defined by the so-called Chinese restaurant process (CRP) as non-parametric priors on groups (Aldous, 1985). To accommodate each answer having multiple categories, we used the multinomial likelihood formulation of the IRM defined in Mørup et al. (2014).

The method relies on Bayesian non-parametric modelling to unsupervised learn coherent prominent structures at the level of groups, inferring from a hypothetical space of an infinite number of groups a suitable representation of groups (in this context, question groups and student groups, respectively) (see also Ghahramani, 2013 and Schmidt & Morup, 2013, for reviews of this methodology). The Multinomial Infinite Relational Model (MIRM) is based on the following generative process (Mørup et al., 2014):

$\boldsymbol{z} \sim CRP(a)$ **Clusters of tasks,**
$\boldsymbol{w} \sim CRP(b)$ **Clusters of students,**

$\boldsymbol{h}_{lm} \sim Dirichlet(\boldsymbol{g})$ **Distribution of answer types of task group l and student group m,**
$A_{ij} \sim Multinomial\left(\boldsymbol{h}_{z_i w_j}\right)$ **Generated response for task i of student j,**

Task group $\boldsymbol{z}$ and student group $\boldsymbol{w}$ are generated according to a CRP defined by assigning observations to groups proportionally to how many are already assigned to the group. Furthermore, these factors are used to define new groups proportional to the parameters a and b respectively. $\boldsymbol{h}_{lm}$ defines the probability of observing the different answer types for the block defined by the l^{th} task group in $\boldsymbol{z}$ and m^{th} student group in $\boldsymbol{w}$. Finally, the generated response A_{ij} to task i of student j is drawn based on this probability distribution $\boldsymbol{h}_{z_i w_j}$ from the corresponding task and student groups.

We used the inference procedure described in Mørup et al. (2014) based on Markov chain Monte Carlo sampling of the parameters, including inferences of the hyperparameters a , b and $\boldsymbol{g}$ by imposing improper priors on these. We treated 1% of the entries in the data as missing and inferred the parameters based on the remaining responses to tasks. These removed answers were then accordingly set to 'unobserved' in **A**. Therefore, tasks that were not answered were treated as missing in the inference procedure. We ran the inference 20 times and reported the sample that obtained the highest joint distribution as defined by

$$\log P(\boldsymbol{A},\boldsymbol{z},\boldsymbol{w}|a\,,b,\boldsymbol{g}) = \left[\prod_{l,m}^{LM}\frac{B(\boldsymbol{r}(l,m))}{B(\boldsymbol{g})}\right]\left[\frac{\Gamma(a)a^{L}}{\Gamma(a+I)}\prod_{l}^{L}\Gamma(\boldsymbol{n}(l))\right]\left[\frac{\Gamma(b)a^{M}}{\Gamma(b+I)}\prod_{m}^{M}\Gamma(\boldsymbol{m}(m))\right]$$

where $r_k(l,m) = \sum_{i:z_i=l,j:w_j=m}^{I,J}\delta(A_{ij},k)$, $\boldsymbol{n}(l) = \sum_i^I \delta(z_i,l)$ and $\boldsymbol{m}(m) = \sum_j^J \delta(w_j,m)$ respectively counts the number of occurrences of answer type k in block (l,m) and size of row cluster l and column cluster m. $B(\boldsymbol{q}) = \frac{\prod_{k=1}^{K}\Gamma(q_k)}{\Gamma(\sum_{k=1}^{K}q_k)}$ is the generalised Beta function.

According to the model, the expected probability of each answer type is given by $\mathrm{E}(\boldsymbol{g}(l,m)) = \frac{\boldsymbol{r}(l,m)+\boldsymbol{g}}{\sum_{k=1}^{K}r_k(l,m)+g_k}$, such that we predicted the most likely answer as $\mathrm{k}^* = \operatorname{argmax}_k \mathrm{E}\,(g_k(l,m))$. For comparison, we included as baselines predictions of the i) most common answer type, ii) most common answer by each student and iii) most common answer for each task. These three baselines were compared to the performance of the MIRM using a McNemar's test based on p-values and confidence intervals as described respectively in Altham (1971) and Bloch and Watson (1967). We further quantified consistencies in the extracted clusters using normalised mutual information (NMI) defined by $NMI(\boldsymbol{q},\boldsymbol{r}) = \frac{2MI(\boldsymbol{q},\boldsymbol{r})}{H(\boldsymbol{q})+H(\boldsymbol{r})}$, in which $MI(\boldsymbol{q},\boldsymbol{r})$ is the mutual information between the partitions $\boldsymbol{q}$ and $\boldsymbol{r}$, whereas $H(\boldsymbol{q})$ and $H(\boldsymbol{r})$ are their respective entropies.

Results of the Analyses

In this section, we reveal the results of running the co-clustering model based on the MIRM. We present the data from the sample that obtained the best solution identified by the model by running 25 initialisations. The best solution was defined as the run that attained the highest $\log P(\boldsymbol{A},\boldsymbol{z},\boldsymbol{w}|a\,,b,\boldsymbol{g})$ for a given iteration. The results of this iteration were then used in the subsequent analysis. The model was run on both the original data set and the subset of students who responded to at least 100 tasks.

The results from running the co-clustering on the data can be visualised in a block structure, as presented in Figure 2. The co-clustering identified seven groups of students, where the last three (5, 6 and 7) were very small. The red colour, which is rather distinguishable and dominant among the four visible groups, represents the correct answers. The horizontal black lines separate the clusters of tasks, with 395 groups identified by the model. These groups of tasks vary in size. The largest cluster consisted of 20 tasks, and many clusters had only a single task in them. We return to the tasks that define these clusters in the discussion.

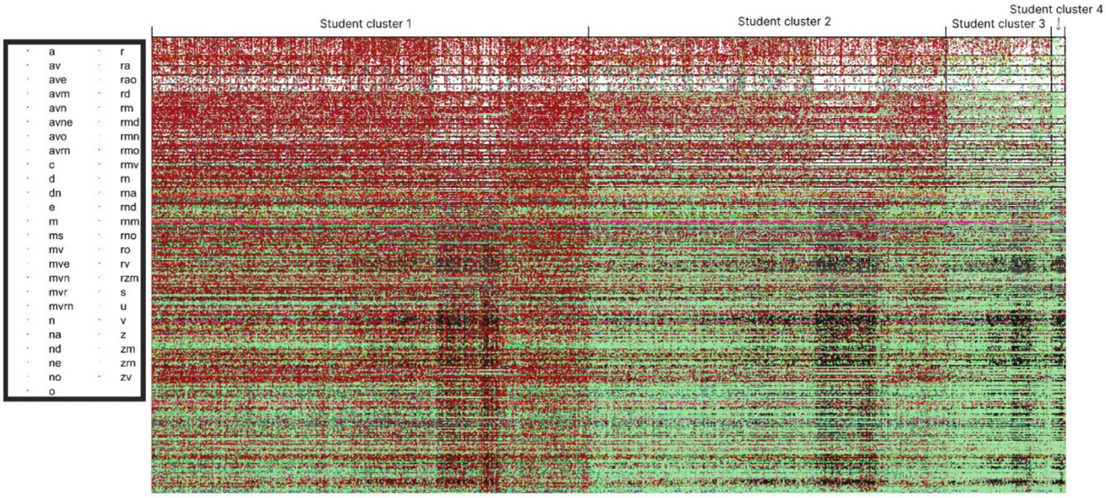


Figure 2: Extracted block structure generated by the Multinomial Infinite Relational Model (MIRM). The model extracted 395 question groups (indicated by rows) and 7 student groups (indicated by columns in which the last three groups of 9, 5, and 2 students are not visible).

To some extent, we observed that the student clusters in Figure 2 correspond to some competence level that allowed them to solve the linear equations. With the red colour representing correct answers, we observed that the number of red dots decreased as we moved Student Clusters 1 to 4.

When observing the structure or grouped data points in Figure 2, we can begin to make sense of the clusters. The more dominant red colour corresponds to the correct answer. Instead of explaining every colour in the figure, we extracted some highlights to discuss the benefits and limitations of observing such student clusters after using the co-clustering method. Figure 2 shows, based on the red colour, that three relevant groups of students (combining the lower four groups into one) seemingly correspond to what we interpret as equation-solving competency levels. More red can be found for Student Group 1, and the remainder of the groups showed a decreasing amount. We remind the reader that the horizontal lines in the figure do not correspond to the types of linear equations the students were able to solve but rather the identified groups of linear equations the groups of students were able to solve, with the largest group of equations highest in the figure. We return to the meaning of the groups of identified tasks in a later section. For clarification, Figure 2 does seem to feature a large amount of what could be interpreted as green dots alongside the red dots. However, the many shades of green in the figure seem to blend easily and can be interpreted as one colour when there are actually several. Every code for the provided answers is present in the figure to display all the codes in their entirety, as the colours are too similar to enable interpretations based solely on Figure 2.

Figure 3 was used to systematically investigate the presence of the different codes or meanings behind the answers. The difference in saturation can be observed as the overall presence of a particular code in the respective student groups. The student groups are signified in columns from left to right, and the groups of tasks are signified as rows. This provides insight into what actually separated the student groups, other than correctness and competency levels. We shall elaborate on this in the Discussion section. The first student cluster consists of 45,114, the second 36,965, the third 10,873, the fourth 1400 and the fifth, sixth and seventh 16 students.

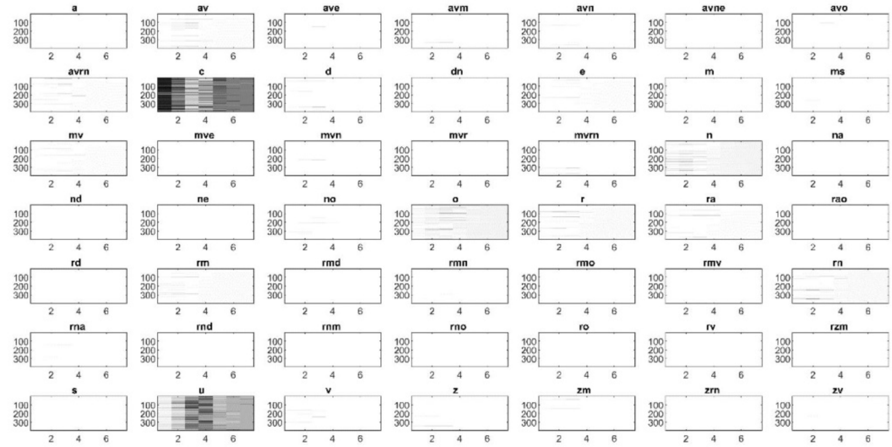


Figure 3: Extracted probabilities $\boldsymbol{h}_{lm}$ of answering each question category for the extracted question groups (rows) and student groups (columns).

When observing the structure in Figure 3, we want to highlight some of the most interesting observations in the identified differences in the student and task groups. Zooming in on the respective error types, we can observe how the given error type is present in the different task and student groups. We took the outset in Table 5 in which we have listed the error types that occurred most often. Observing the structures for the correct and unknown answers, we noticed that the first group had the highest number of correct answers. This can be observed through the saturation of the black lines in each column. When observing the correct and the unknown codes at the same time, we can see that they were almost opposite each other (observing Groups 1–4). We shall return to the presence of unknown answers when discussing the respective groups of tasks.

Table 5: Overview of the code frequency

Name of the Code	**Tag**	**Count**
Correct	c	1,722,308
Unknown reason	u	182,450
Negativity issue	n	73,373
Rearranging for sense-making	r	36,246
One more/one less	o	36,046
Rearranging and negativity issue	rn	27,863
Rearranging and multiplication issue	rm	13,417
Adding the coefficient to the unknown	av	9,098
Ignoring the coefficient	mv	6,279
Equals sign	e	5,511

Through the most common error (if we observed them in isolation and disregard correct 'c' and unknown reason 'u'), we can observe the structure of the 'negativity issue' in Figure 4. As mentioned, this code was provided to answers where the additive inverse of the correct answer was chosen. We know from earlier studies that such errors are common in lower secondary school (Gallardo, 2002; Vlassis, 2002).

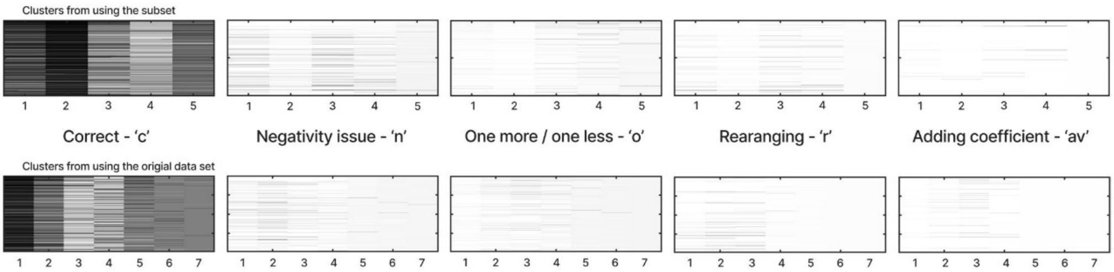


Figure 4: Highlights of extracted probabilities, both from the model on the original data set and the model on the subset

Notably, the blacker the lines in the column or the more saturated the column is, the more this type of error occurred. In addition, the black lines each signify a row that corresponds to a group of tasks. In this section, we do not dive into the concrete group of tasks that forced these errors. The 'negativity issue' is highly represented in most of the equation types. However, this issue in particular did not become increasingly present going through the student groups. In the following, we comment on the highlights from the cluster in the original data set presented in Figure 4. The 'negativity issue' was increasingly present in the second group of students. However, the third group seemed to make fewer of these errors compared to the second group. Comparing the first and third student groups, this error occurred for different groups of tasks. However, this does not mean that the tasks were different. Instead, the error was perhaps equally present but maybe related to different tasks that the clustering algorithm perceived as different tasks for some other reason.

The error we identified as 'rearranging for sense-making' yielded a different pattern. This reason for these errors stems from the fact that some students were not able to make sense of the expression or equations and therefore rearranged the entities within them to be able to provide an answer. This error mostly occurred in the third type (Type 3) of equations (e.g., $3 - x = 6$), where we observed that students were more willing to provide 9 as an answer than they would the correct answer -3. Reading the expression as $x - 3 = 6$ allowed the students to arrive at the solution $x = 9$. We could mostly observe that the error persisted throughout the respective student groups; however, Group 4 was not as likely to make this error.

Next, we assessed the prevalence of the 'one more/one less' error type (see Figure 6). This error was particularly interesting to observe when coding the data during the data enhancement process. The error is simple, in that the students hit the wrong key on the keyboard when putting in the answers in the environment. We observed that this error was a common error among all types of equations, which, in many ways, would make sense. One could intuitively imagine that a similar error would not exist on pen and paper, or at least not for the same reasons that we believe that the error exists in digital learning environments.

We do not believe this error corresponds to students' equation-solving abilities; rather, it boils down to some level of concentration or motivation. We imagine that a student wanting to finish the tasks faster would be more likely to accidently press the wrong button. Observing the progression in this error type through the groups of students, we could see that the frequency of this type of error occurring was relatively the same for all groups. This means that only approximately half of the task groups provoked this issue.

Going through these error types in the data that were most prominent, we want to mention issues that were present but not at all surprising to find. Kieran (1985) mentioned in their list of errors students make when solving equations that understanding the concatenation of numbers and letters leads to students adding the coefficient on to the unknown instead of multiplying them. The first error type we want to emphasise is 'av', which refers to the issue where students reach a solution where the coefficient is added onto the unknown. The first thing we noticed was that the error was almost absent in the first student group. Second, the error also seemed to be most prominent among this student group for a selected group of tasks. We shall discuss the actual tasks in the following section, in which we dive into the task specifics. As one might have expected, the issue becomes more prevalent throughout the student groups. This same pattern was true for the error type 'mv', which is also connected to issues with the concatenation of numbers and letters. However, this issue occurred when students disregarded the coefficient and treated the unknown and the coefficient as one entity to solve the equation.

Another way to extract more interpretable results from the model was to look at the most prominent errors within each student group. The first student cluster consisted of high-performing students in terms of equation-solving ability. Their table of distributions of codes shows the following:

Group 1	c	u	n	o	r	rn	rm	Distributed answers	Total answers
Answers	1045468	41900	29237	9933	10310	10689	5644	1153181	1165441
%	89.7058	3.5952	2.509	0.852	0.885	0.917	0.484	98.948	100

This high-performing group delivered around 90% correct answers to the equations. The most common error, when disregarding their answers for which we could not justify a reason, was answers connected to providing the additive inverse as an answer to the tasks. Looking at the behaviour of Group 1, we observed that approximately 7% of their answers were interpretable errors, and these were mostly errors connected to providing the additive inverse as the answer (negativity issue 'n'). This group, in most cases, provided incorrect answers as a result of this issue, based on the number of errors relative to the total number of errors within the respective types of equations. For example, we observed that the three best equations for this group to work with relative to issue 'n' are:

1. -4 – x = 9: 2.23
2. 3x = -15: 2.02
3. 5x = -25: 1.38

Beside each equation is what we define as a relevance score. The score was calculated as the total number of issue 'n' errors made by Student Group 1 divided by the total number of errors for each of their respective types of equations (as shown in Table 2). For example, for equation $-4 - x = 9$, a Type 3 equation, we found 1125 errors with the code 'n'. Regarding the correctness rate, of 79%, Type 3 equations received 50401 erroneous answers. We calculated the relevance score by dividing the number of answers coded with 'n' by the total number of answers (and multiplying by 100, giving percentages for readability). If we performed the same experiment for Student Group 2, we see that this group also revealed the issue of providing an answer that is the additive inverse. However, we did find a different distribution for their answers:

Group 2	c	u	n	o	r	rn	rm	av	Distributed answers	Total answers
Answers	592121	85145	38596	18361	20801	14622	6449	4495	780590	799272
%	74.083	10.653	4.829	2.297	2.603	1.829	0.807	0.562	97.663	100

This group had a significantly lower number of correct answers compared to Group 1. We observed a large increase in the 'unknown reason' and a huge increase in the 'one up/one down' code. This indicates that significantly more of the answers were either uninterpretable, unmotivated or unlucky. Furthermore, we observed a slightly different top three, but not significantly, of equations causing the 'negativity issue':

1. x – 8 = -3: 3.66
2. 3x = -15: 3.10
3. 5x = -25: 2.93

We also observed higher relevancy scores, indicating that these equations received a larger number of erroneous answers than with Student Group 1. We continued to look at the issue ‘r’ (rearranging for sense-making) with Student Group 2, since this issue also caused a large number of errors. Here, we observed the following three equations as the top scorers:

1. 3 – x = 6: 4.05
2. 4 – x = 5: 3.51
3. 6 – x = 3: 3.03

These equations are all Type 3 equations, where x is subtracted from a constant. As indicated earlier, the issue is that these equations cause students to add the number on the left side to the number on the right, neglecting the minus sign. Our hypothesis is that some students might actually have read these equations as $x - 3 = 6$, resulting in their providing 9 as the answer.

Regarding the ‘u’ (unknown reason) and ‘o’ (one more/one less) issues, we observed a clear tendency when progressing down through the groups that these issues increased in prevalence dramatically, especially the ‘u’ issue. This indicates that these groups cannot simply be separated by performance or competency; they may also be separated by concentration or motivation. Normal statistical tools in digital learning environments would simply interpret this behaviour as incorrect and would signal to the teacher that students from the lower groups are simply providing incorrect answers. With the added coding system, we can provide feedback implicating concentration issues, since the answers would be tagged with ‘unknown reason’. In Student Group 3, we needed to observe even more codes in the significant distribution of answers:

Group 3	c	u	n	o	r	rn	rm	av	mv	ra	e	Distributed answers	Total answers
Answers	82489	46698	5255	7016	4531	2477	1255	2811	1174	848	883	155437	158118
%	52.169	29.534	3.324	4.437	2.866	1.567	0.794	1.778	0.743	0.536	0.560	98.304	100

With Group 3, the proportion of correct answers fell to 52%. We observed a clear separation in the groups according to the number of correct answers they provided. Notably, the number of unknown reasons for putting in a wrong answer also dramatically increased through the groups.

Group 4	c	u	n	o	rn	rm	av	Distributed answers	Total answers
Answers	2257	8524	273	721	75	68	230	12148	12518
%	18.030	68.094	2.181	5.760	0.599	0.543	1.837	97.044	100

With Group 4, the proportion of correct answers fell to below 20%, and the ratio of unknown reasons was almost 70%. This indicates that even though these groups were co-clustered according to equation-solving skills and the number of correct answers provided, other factors linked these groups. We observed that these first four identified groups of students might seem to face similar challenges. However, if we dive into the specifics, we find that something else is present in the data. The following table summarises the major differences between the four prominent student groups:

Student Group	Correct	Unknown Reason	Sum
1 - 45,114	90%	4%	94%

2 - 36,965	74%	11%	85%
3 - 10873	52%	30%	82%
4 - 1400	18%	68%	86%

Only 6% of Group 1's answers were errors that we can identify the reasons for these errors and base assessment reports on. Almost a third of the errors involved a negativity issue, where the additive inverse of the actual solution was chosen as the answer. The later groups more consistently left about 15% of their answers as errors to which we could apply a code signifying a reason. The same issues for the top scorers occurred for every cluster of students.

In Table 6, we present the results of how well the MIRM performed against the three common benchmarks. This result was measured against the 1% of the original data left out of the model in each of the 25 models, from which we chose the best. The results show that the MIRM performed better than the three standard benchmarks.

Table 6: Answer prediction results leaving 1% of answers as missing during inference predicted by i) overall most common answer type, ii) most common student answer, iii) most common question answer and iv) most probable answer according to the MIRM. According to McNemar's test, the MIRM significantly outperformed the other approaches ($p < 10^{-12}$) with the smallest effect size being between the most common question answer and MIRM (95% confidence interval of difference in accuracy [0.72%; 1.26%]).

Most common answer	**Most common student answer**	**Most common question answer**	**Most probable MIRM answer**
80.59% accuracy	80.61% accuracy	80.81% accuracy	81.80% accuracy

We further quantified the reliability across the 25 trained models using NMI by comparing the extracted group structure of the first run to the second, second to the third, etc., such that the last run was compared to the first run. Whereas the identified task groups had an average NMI of 0.941+/-0.002, the extracted student groups were not found to be reliable and had an average NMI of 0.139 +/- 0.003 across the 25 runs, indicating substantial differences across the reruns in how students were assigned to groups. We attribute this to the fact that for students with few responses, it was hard to determine to which group they should be assigned. This is supported by the corresponding analysis of the reduced subset of the data that contained only students with at least 100 responses, for which we found substantially higher consistency between the reruns as measured by NMI (see the Appendix).

Regarding the groups of tasks, we extracted the information on the tasks that were able to perform similarly enough that they would group (clustered with the co-clustering model). In the following, we present the different clusters of tasks and the types of equations present within them. The clusters we found among the tasks were quite small and many in quantity. The biggest cluster consisted of 20 tasks, and the smallest cluster contained only a single task. The tasks are distributed in the clusters as presented in Figure 5.

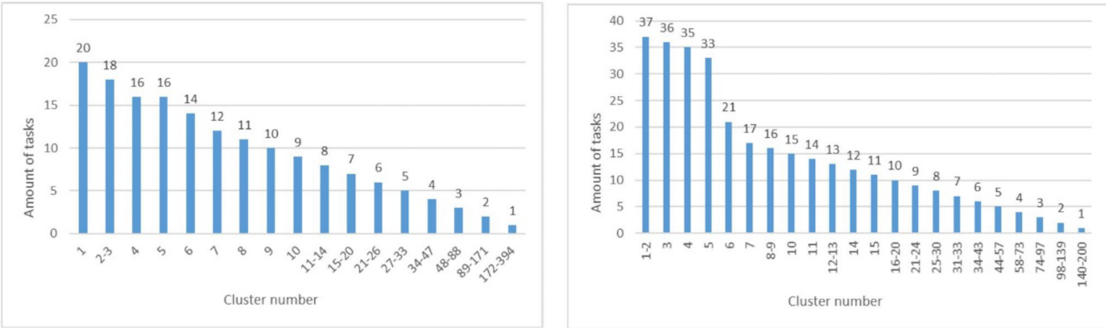


Figure 5: Task distribution across clusters in both data sets (original on the left, subset on the right).

With the outset in these clusters, we present in the following the types of equations that were actually grouped together by the co-clustering algorithm for presenting similar challenges to students based on the coding of their answers. Identifying the behaviours that result in the groups presented in Figure 5 makes it possible to determine which tasks provoke a certain type of behaviour or error when students engage with them.

Table 7: The tasks and dominating error codes represented in the groups of tasks based on the full data set

Cluster number	**Present types of equations**	**Examples of types**	**Dominating error type**
1 – 36 tasks	83% of Type 3 rev 17% of Type 3	9 = 13 - x 25 -x = 10	R, n
1 - 20 tasks	100% of Type 7 rev	8 = 2x + 6, 12 = 3x + 6	Av, mv
2 - 18 tasks	100% of Type 7 rev	7 = 2x – 1, 8 = 2x – 4	Rn, e
3 - 18 tasks	89% of Non-simplified Arithmetic 11% of Type 7	3x – 7 + x = 13, 6x – 9 = 21	Rn, avrn, o
4 - 16 tasks	100% of Non-simplified Arithmetic	5x – 3 - 3x = 7	O, rn
5 - 15 tasks	80% of Type 2 20% of Type 3	X - 14 = 16, 5 = 8 – x	R, o
6 - 14 tasks	43% of Type 7 57% of Type 7 rev	8 = 3x + 2, 7x + 27 = 41	O, av, mv
7 - 12 tasks	100% of Type 1	7 + x = 9	O (many correct answers)
8 - 11 tasks	6/11 of Type 7 4/11 of Type 4 1/11 of Type 7 rev	67x + 13 = 147, 3x = 18, 14 = 3x - 4	O, av
9 - 10 tasks	50% Type 7 50% Type 7 rev	5x + 12 = 32, 26 = 5x + 6	O
10 - 9 tasks	8/9 of Type 3 1/9 of Type 2	15 – x = 11, x – 5 = 9	R, o, n

Observing the differences between the dominating error types in Clusters 1 and 2 calls for an in-depth analysis of the concrete tasks and their differences in construction. We observed that among many of the tasks that were able to be grouped together, the larger groups of tasks were very much dominated by Type 7 equations ($ax + b = c$) and their reversed variations ($c = ax + b$). Revisiting the frequency table for the specific error codes, we attempted to locate and discuss the particular groups of equations that best determine a specific error type.

Table 6: Examples of tasks attracting most erroneous answers related to the most frequent error types

Name	Tag	Best tasks for detection of issue
Unknown reason	u	42/(x + 14) = 4, 3x – 7 = 6x + 2
Negativity issue	n	x – 7 = -4, 5x = -25, -7x + 30 = 9, -4 – x = 9, x – 5 = -2
Rearranging for sense-making	r	3 – x = 6, 4 – x = 5, 6 – x = 3, x – 43 = 77, 4 = 13 – x, x – 7 = 15
One more/one less	o	4 + x = 9, 9 + x = 15, 7x + 14 = 63
Rearranging and negativity issue	rn	2x + 1 = x – 4, x – 4 = 2x + 1, 8x + 56 = 5x + 20, x - 4 + 3x = 8
Rearranging and multiplication issue	rm	x/2 = 2, x/5 = 5, x/5 = 20, 4/x + 3 = 7
Adding the coefficient to the unknown	av	4x + 3 - 2x = 13, 8/x – 8 = -4
Ignoring the coefficient	mv	5/x – 6 = -5, 33/x – 33 = -22

Observing the clusters of tasks found in the subset, we see that sorting the clusters first on percent-wise distribution of the answers among the error codes and then sorting by the highest value of the 'negativity issue' allows an opportunity to locate tasks particularly good at provoking certain errors (see Table 9).

Table 7: Subsection of task cluster sorted to locate the clusters that are most likely to cause a negativity issue

Task cluster number	c	u	n	o	r	Total%
C 93	49	6	45	0	0	100
C 94	49	16	33	0	0	98
C 185	66	5	27	0	1	99
C 200	60	8	26	6	0	100
C 88	65	7	23	0	0	95
C 141	53	19	23	0	4	99
C 120	73	5	21	1	0	90

Taking note of the distribution of answers within the clusters lets us observe what we can characterise as diagnostic value in this context. Looking at the tasks from Task Cluster 93 (the first row in Table 6), we see that 45% of the answers provided to these tasks were incorrect due to the 'negativity issue'. Equally important is that the remaining answers were distributed among very few other codes. In this case, there were 49% 'correct' and 6% 'unknown reason' answers. The particular tasks in Cluster 93 (of the subset) are:

- $-3x = -15$ (Type 4)
- $-4 - x = 9$ (Type 3)
- $12x + 57 = 2x + 27$ (type 'General algebraic')

With our results presented, we now dive into the discussion of our posed research question. We have presented the results of running the MIRM on a data set (and a subset) that represents the behaviours of students working in a digital learning environment. Our data set was extracted and managed with the intention of observing something that is close to what actually happens in a Danish school when students solve equations on Denmark's largest digital resource for mathematics. Our research question was as follows:

How can the method of co-clustering contribute to the analysis of a large partially observed data set collected in a digital learning environment to generate better possibilities for learning about task design and students' difficulties solving linear equations?

We managed to identify groups of students and groups of tasks and extracted the characteristics of these referring to Hypotheses a) and b). The signal in the enhanced data set was strong enough that the MIRM was able to form co-clusters. Furthermore, the MIRM performed better than the three baselines we compared it to when predicting students' responses to tasks treated as unobserved. The data that we believed would be the easiest to extract from digital learning environments were data sets, such as the one we have presented the analysis of in this paper. We realise that one could, of course, limit the data set to observations of a group of students that would satisfy the criteria of a fully observed data set, as we have done with the subset. However, we believe that with the MIRM, we create new opportunities for data analysis that can lead to improved and faster assessment for task designers for digital learning environments. We showed that co-clustering models, such as the MIRM, can meaningfully complete the data and make clustering analyses possible. In fact, the MIRM estimated the missing answers in the data set with more accuracy than the three baselines shown in Table 6. This in itself helps in determining the relevancy of tasks in teaching in the way that groups of tasks can help verify design principles. When tasks attract similar and distinct behaviours, we can begin to make new assumptions about tasks grouped together in learning to exemplify certain difficulties across what might seem to be very different equations.

Regarding the first four groups of students, the clustered data indicate what kinds of difficulties are represented by the coded errors each group experiences. From the results of running the model, we observed that the NMI when forming the student groups was low, telling us that the placing of students into clusters was very different across the 25 models made. This means that we should not make too many assumptions based on what group any individual student is assigned. We did, however, observe a much stronger signal in the subset, with a much higher NMI assigning the students, reassuring us that the clusters found in the best model using the original data set were not rejected, since the clusters found using the subset were quite similar in structure. The NMI assigning the tasks into clusters was very high in both models, indicating that the tasks were stable in their extracted group structures and that the signal in the answers they received was strong. In the clusters found by running the model on the subset, we obtained many fewer clusters, making them more interesting from a math education perspective. The structure of these clusters of tasks is an excellent tool for verifying the anticipated performance for a task, especially when designed as suggested by Elkjær and Jankvist (2021). The idea of using variation theory to design a comprehensive set of tasks covering the identified difficulties students face when solving linear equations is based on the idea that these tasks should do so. However, to verify the anticipated performance of such a design, we suggest co-clustering with a data set representing 'normal behaviour' found in a digital learning environment.

We observed a clear pattern with the error types. Even though the 'unknown reason' was usually the top scorer after the correct answer, this was expected. The 'negativity issue' might also be expected to be present, as we observed in our data analysis. What is different and perhaps an important realisation is the presence of the 'one more/one less' issue. The number of answers that received this code and its presence in task clusters makes it a significant consideration when designing digital learning environments. When interpreting students' equation-solving abilities, we observed a solid reason to go beyond correctness in an endeavour to interpret students' errors.

In relation to the validation of task design, however, the 'one more/one less' and 'unknown reason' issues contributed in a different manner. The verification of the anticipated performance of the tasks relies on these parameters in a different way. Assuming that the presence of these two issues in a cluster is high, one might argue that the tasks in such a cluster are difficult to justify as being effective in highlighting a certain type of error. However, we identified that clusters with a low number of error types and a high frequency of one specific error exist. Such clusters, and in particular the tasks within them, are important to observe, since these tasks are likely to demonstrate a very specific type of performance.

When coding the five most popular answers for each of the 892 tasks, we realised that some codes were almost absent in the data at all. This was not a consideration while coding, since the coding enabled the model to co-cluster on something arguably better than the actual answers or if the answers were correct or not. However, by observing the overall distribution of the codes within each cluster, we argue that some of the codes could benefit from being more nuanced. The code 'r'—rearranging for sense-making—might evolve into something more specific, since 'rn' (rearranging and a negativity issue) and 'rm' (rearranging and a multiplication issue) are not dominant. Through Kieran (1985), we were familiarised with students inversing subtraction with subtraction and addition with addition when transforming equations to reach a solution. Interpreting the answers from the digital learning environment, we were not able to apply codes matching Kieran's observations exactly, since we could not be sure in determining the students' trains of thought when solving a task such as $3 - x = 6'$. Some of the 49 codes were extremely rarely represented, and the model might benefit from a sorted data set where some of the rarely represented answers are filtered out.

Co-clustering is notoriously a challenging NP-hard problem, with no guarantee of identifying the optimal solution. We considered 25 reruns and selected the sample with the highest joint distribution to mitigate local minima issues. We further quantified the consistencies across these reruns using the NMI. However, we found that although the identified task groups appeared to recover reliably across runs, the identified student groups differed substantially. Care should therefore be taken when interpreting the co-clustering results, as different obtained solutions may result in changing interpretations. We presently used non-parametric Bayesian modelling to identify suitable numbers of row and column clusters. Although this approach enables efficient model complexity, exploration is not guaranteed to identify the correct model order.

Conclusion

We discussed in which ways the method of co-clustering can contribute to the analysis of a large partially observed data set collected in a digital learning environment to generate better possibilities for learning about task design and students' difficulties when solving linear equations. We have presented and discussed several aspects of the data analysis and in which ways the co-clustering can contribute to the determination of student and task groups that might yield positive insights for knowledge on and the verification of task design for digital learning environments.

The hypothesis was that prominent groups of students who systematically face similar challenges across groups of linear equations could be studied through partially observed collective information from a digital learning environment. To guide the taxonomy towards exploring a space in which task designers would benefit from receiving verification on students' anticipated performance for a given task, we went with the idea of applying the reasons for specific student answers. This meant that we, to some extent, had to reverse-engineer the idea behind the design of the 892 tasks. The tasks were initially designed with the assumed ability to exemplify or provoke certain more or less classical errors. In a retrospective review of the coding system (see Table 2), we realised that much of the anticipated behaviour was captured in the data (see Table 5). The five most popular answers presented a little over one unidentifiable or unknown reason for an answer on average. The number of answers provided without an identifiable reason also seemed to rise with the complexity or difficulty of the linear equations. In addition, exploring the popular answers and coding them with reasons for the particular answer being chosen enables not only a multiple-choice version of the task used in this analysis but also possibilities for new assessment strategies in digital learning environments. If implemented in certain areas of the digital content, teachers could acquire a different kind of statistical overview incorporating the reasons for the answers and thereby information that could be tracked—in our case, this includes different types of linear equations—but also possibly extended over several mathematical subject areas.

The grouping of students, along with the insight into what type of errors the students in the respective groups actually struggle with, can benefit teaching. We observed that even students whose answers were almost 90% correct made errors related to confusing the correct answer with the additive inverse. We observed a clear sign that this particular negativity issue was prominent among the other groups as well. Issues tied to the equals sign were not particularly present in our results. Here, we would like to emphasise that the observed code 'r' (rearranging for sense-making) could be related to issues connected to the equals sign. We suspect that these errors are more connected to procedural or perceptual issues when students interpret the task before them and apply a solution strategy. Importantly, we noticed that the 'one more/one less' issue was among the top scoring reasons for errors in this data set. In many ways, such answers could be characterised as correct, and the reason we must apply a different code is not cognitive. We consider this an important finding, since it is only when going beyond correctness that a digital learning environment can begin to make better sense of these errors.

Acknowledgements

This work is possible due to grants from Innovation Fund Denmark, grant 8053-00042A

On behalf of all authors, the corresponding author states that there is no conflict of interest.

The datasets generated during and/or analysed during the current study are not publicly available due to GDPR.

References

Aldous, D. J. (1985). Exchangeability and related topics. In *École d'Été de Probabilités de Saint-Flour XIII—1983* (pp. 1–198). Springer.

Altham, P. M. (1971). The analysis of matched proportions. *Biometrika*, *58*(3), 561–576.

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability (Formerly: Journal of Personnel Evaluation in Education)*, *21*(1), 5. https://doi.org/10.1007/s11092-008-9068-5

Bloch, D. A., & Watson, G. S. (1967). A Bayesian study of the multinomial distribution. *The Annals of Mathematical Statistics*, 1423–1435.

Bokhove, C. (2017). Supporting variation in task design through the use of technology. In *Digital Technologies in Designing Mathematics Education Tasks* (pp. 239–257). Springer.

Elkjær, M., & Jankvist, U. T. (2021). Designing tasks for a dynamic online environment: Applying research into students' difficulties with linear equations. *Mathematics*, *9*(5), 557–557. https://doi.org/10.3390/math9050557

Filloy, E., & Rojano, T. (1989). Solving equations: The transition from arithmetic to algebra. *For the Learning of Mathematics*, *9*(2), 19–25.

Gallardo, A. (2002). The extension of the natural-number domain to the integers in the transition from arithmetic to algebra. *Educational Studies in Mathematics*, *49*(2), 171–192.

Ghahramani, Z. (2013). Bayesian non-parametrics and the probabilistic approach to modelling. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *371*(1984), 20110553.

Govaert, G., & Nadif, M. (2013). *Co-clustering: Models, algorithms and applications*. John Wiley & Sons.

Holland, P. W., Laskey, K. B., & Leinhardt, S. (1983). Stochastic blockmodels: First steps. *Social Networks*, *5*(2), 109–137.

Jankvist, U. T., & Niss, M. (2015). A framework for designing a research-based "maths counsellor" teacher programme. *Educational Studies in Mathematics*, *90*(3), 259–284. https://doi.org/10.1007/s10649-015-9629-8

Kemp, C., Tenenbaum, J. B., Griffiths, T. L., Yamada, T., & Ueda, N. (2006). *Learning systems of concepts with an infinite relational model*. *3*, 5.

Kieran, C. (1981). Concepts associated with the equality symbol. *Educational Studies in Mathematics*, *12*(3), 317–326. https://doi.org/10.1007/BF00311062

Kieran, C. (1985). The equation-solving errors of novice and intermediate algebra students. In L. Streefland (Ed.), *Proceedings of the Annual Conference of the International Group for the Psychology of Mathematics Education* (pp. 141–146).

Küchemann, D. (1981). Algebra. In K. Hart (Ed.), *Children's understanding of mathematics: 11–16* (pp. 102–119). Murray.

Leung, A., & Baccaglini-Frank, A. (2017). Digital technologies in designing mathematics education tasks. *Mathematics Education in the Digital Era*.

Linsell, C. (2009). A hierarchy of strategies for solving linear equations. In R. Hunter, B. Bicknell, & T. Burgess (Eds.), *Crossing divides: Proceedings of the 32nd annual conference of the Mathematics Education Research Group of Australasia* (Vol. 1, pp. 331–338). MERGA.

Marton, F. (2015). *Necessary conditions of learning*. Routledge.

Matthews, P., Rittle-Johnson, B., McEldoon, K., & Taylor, R. (2012). Measure for measure: What combining diverse measures reveals about children's understanding of the equal sign as an indicator of mathematical equality. *Journal for Research in Mathematics Education*, *43*(3), 316–316. https://doi.org/10.5951/jresematheduc.43.3.0316

Mørup, M., Glückstad, F. K., Herlau, T., & Schmidt, M. N. (2014). *Nonparametric statistical structuring of knowledge systems using binary feature matches*. 1–6.

Nowicki, K., & Snijders, T. A. B. (2001). Estimation and prediction for stochastic block structures. *Journal of the American Statistical Association*, *96*(455), 1077–1087.

Prediger, S. (2010). How to develop mathematics-for-teaching and for understanding: The case of meanings of the equal sign. *Journal of Mathematics Teacher Education*, *13*(1), 73–93. https://doi.org/10.1007/s10857-009-9119-y

Schmidt, M. N., & Morup, M. (2013). Nonparametric Bayesian modeling of complex networks: An introduction. *IEEE Signal Processing Magazine*, *30*(3), 110–128.

Steenbergen-Hu, S., & Cooper, H. (2013). A meta-analysis of the effectiveness of intelligent tutoring systems on K–12 students' mathematical learning. *Journal of Educational Psychology*, *105*(4), 970. https://doi.org/10.1037/a0032447

Tall, D., & Vinner, S. (1981). Concept image and concept definition in mathematics with particular reference to limits and continuity. *Educational Studies in Mathematics*, *12*(2), 151–169. https://doi.org/10.1007/BF00305619

Vlassis, J. (2002). The balance model: Hindrance or support for the solving of linear equations with one unknown. *Educational Studies in Mathematics*, *49*(3), 341–359. https://doi.org/10.1023/A:1020229023965

Xu, Z., Tresp, V., Yu, K., & Kriegel, H.-P. (2006). Learning infinite hidden relational models. *Uncertainty in Artificial Intelligence (UAI2006)*, 2.

Appendix - Supplementary information

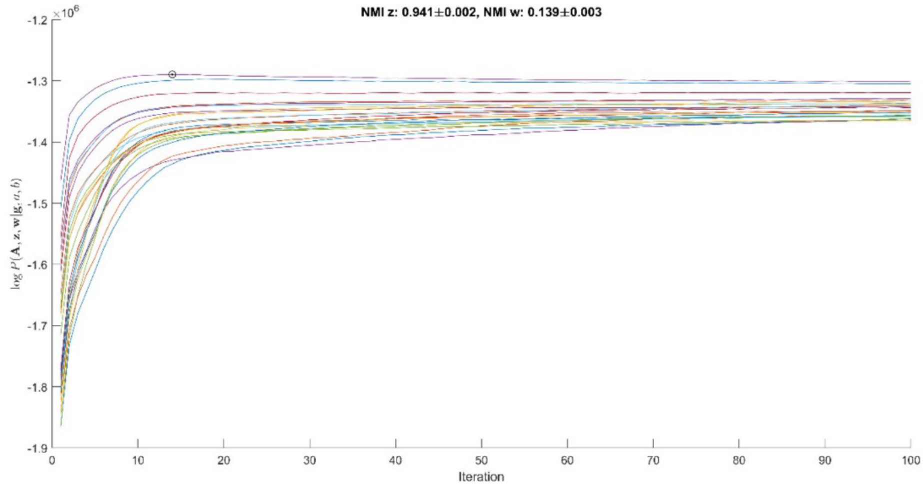


Figure 1s: Model convergence across 25 randomly initialised model and model consistency quantified using normalised mutual information (NMI) for the extracted question groups (NMI Z1) and student groups (NMI Z2). The standard deviation of the mean NMI across the 25 initialisations are indicated by mean ± SD. The best identified solution across the 25 random initialisation and 100 inference iterations is indicated by ⊙.

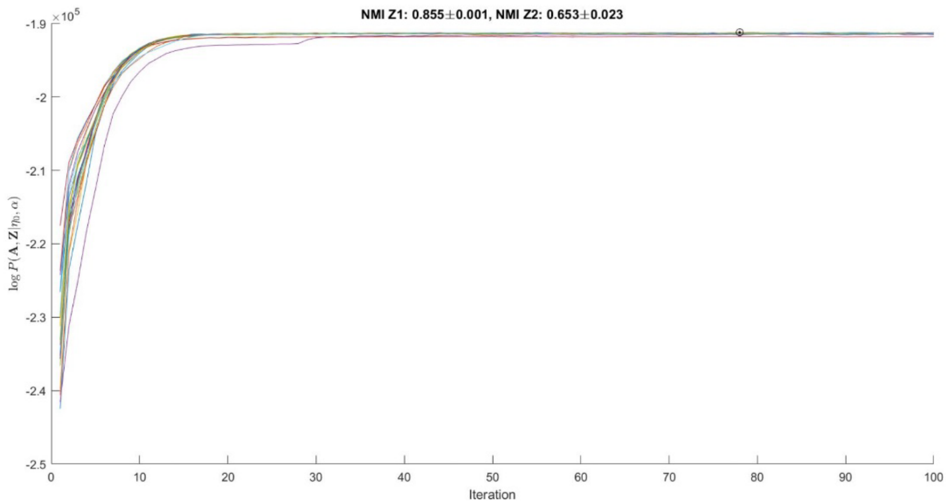


Figure 2s: Model convergence across 25 randomly initialised model and model consistency quantified using NMI for the extracted question groups (NMI Z1) and student groups (NMI Z2). The standard deviation of the mean NMI across the 20 initialisations indicated by mean ± SD. The best identified solution across the 25 random initialisation and 100 inference iterations is indicated by ⊙.

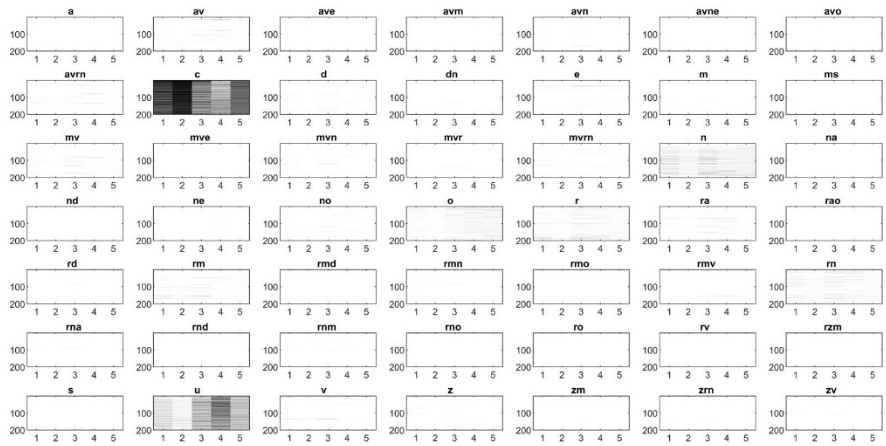


Figure 3s: Extracted probabilities h_lm of answering each question category for the extracted question groups (rows) and student groups (columns) using the subset.

Paper F

Elkjær, M., & Thomsen, L. A. (2022). Adapting the balance model for equation solving to virtual reality. *Digital Experiences in Mathematics Education*. Springer https://doi.org/10.1007/s40751-022-00103-4

Please note that this published paper contains errors. Two direct quotes from Vergnaud (2009) are missing from the section labelled 'The Theory of Conceptual Fields and the Notion of Scheme' on page 129 after each of the colons. The quotes are listed here:

> It has two aims: (1) to describe and analyse the progressive complexity, on a long- and medium-term basis, of the mathematical competencies that students develop inside and outside school, and (2) to establish better connections between the operational form of knowledge, which consists in action in the physical and social world, and the predicative form of knowledge, which consists in the linguistic and symbolic expressions of this knowledge. As it deals with the progressive complexity of knowledge, the conceptual field framework is also useful to help teachers organize didactic situations and interventions, depending on both the epistemology of mathematics and a better understanding of the conceptualizing process of students. (Vergnaud, 2009, p. 83)

> [Schemes] describe ordinary ways of doing, for situations already mastered, and give hints on how to tackle new situations. Schemes are adaptable resources: they assimilate new situations by accommodating to them. Therefore, the definition of schemes must contain ready-made rules, tricks and procedures that have been shaped by already mastered situations. (Vergnaud, 2009, p. 88)

Digital Experiences in Mathematics Education (2022) 8:127–156
https://doi.org/10.1007/s40751-022-00103-4

Adapting the Balance Model for Equation Solving to Virtual Reality

Morten Elkjær[1] · **Lui A. Thomsen**[2]

Accepted: 10 March 2022 / Published online: 7 April 2022

Abstract
This article presents the theoretical considerations leading to the design and development of a digital experience for teaching linear equations using a modified balance model for equation solving. We modified the balance to alter physics behaviour in a virtual reality (VR) experience, to strengthen students' schemes for solving linear equations and help students to adapt their schemes to situations where negative numbers and mathematical negativity make equations abstract. We used the VR application in a small teaching experience with ten students and their mathematics teacher from a Danish grade 7 class (13–14 years of age). The exploratory study aimed to analyse and evaluate the effect of teaching with the modified balance in the VR application via a novel teaching experience. We report findings that show positive prospects for the use of VR in teaching linear equation solving including a new equation solving strategy enabled by the virtual environment. A majority of students gave a positive affective response to the experience, referred to, and were able to apply ideas from the VR experience to linear equation solving exercises on post-experience pen-and-paper exercises. Moreover, we report findings from a particular student case who showed interesting behaviour and reasoning, from which we provide in-depth analysis to understand future possibilities of teaching equation solving with VR.

Keywords Virtual reality · Linear equations · Negative numbers · Strategies · Balance model

✉ Morten Elkjær
me@edu.au.dk

Lui A. Thomsen
lat@create.aau.dk

1 Aarhus University, Danish School of Education, Copenhagen, Denmark

2 Aalborg University, CREATE, Multisensory Experience Lab, Copenhagen, Danmark

Vlassis (2002) asks if we should reject the balance model when teaching equation solving. Traditionally, the balance model has been a common tool for teaching and discussing the concept of linear equations for algebra beginners in lower secondary school (Otten, van den Heuvel-Panhuizen & Veldhuis, 2019; Pirie & Martin, 1997; Rhine, Harrington & Starr, 2018). However, the balance model has some difficulties in representing the sometimes-necessary algebra and, therefore, it has limitations as a tool for teaching linear equations.

Several studies show that the balance model has severe shortcomings in representing and facilitating learners' work with negative numbers (Pirie & Martin, 1997; Vlassis, 2002). We believe that emerging technologies, such as virtual reality (VR), may have the potential to facilitate learning in new engaging ways, for example by overcoming limitations of the physical world. Therefore, the current study presents the design of a VR experience for teaching linear equations and linear equation solving strategies involving negativity for use in lower secondary school.

VR, in this context, is composed of a head-mounted display (HMD) and hand-held controllers with positional and rotational motion tracking. Research in the field has mostly been conducted with non-immersive media, such as web-based applets. Research on the use of VR in this context is sparse, but promising, due to the interactive capabilities of VR which is the determining component of virtual manipulatives as a learning and teaching tool (Moyer-Packenham & Bolyard, 2016).

We introduce the notion of a scheme and the theory of conceptual fields (Vergnaud, 2009) as a theoretical lens through which we intend to discuss the understanding of the concept of linear equations and competencies related to solving linear equations. We expand the role of this theory in the theoretical constructs.

We intend to maintain identified affordances of the classical balance model and virtual manipulatives to create a tool for teaching linear equations (Otten et al., 2019). The motivation stems from an overall purpose of designing a tool for teaching linear equations that affords not only the representation of necessary elements (i.e. mathematical equality), but also allows for teaching more advanced equation solving strategies than what the classical balance model permits. To allow negativity to be represented on the balance model, we intend to utilise the affordances of virtual environments and VR. Vlassis (2002) finds that equations such as $-6x = 24$ are extremely difficult for lower secondary school students to make sense of. Therefore, the VR environment should allow for a balance model capable of representing such equations. Linsell (2009) finds that having access to a larger variety of equation solving strategies (rules) strengthens students' schemes related to the concept of linear equations.

The specific research questions related to the overall research aims are:

How may the modified balance model in a VR environment strengthen thestudents" schemes related to linear equations involving negativity?

In addition, how may the built-in 'invert mechanic' further assist the students in developing strategies for solving linear equations involving negativity?

We address the above questions through a teaching experience with a fully functional implementation of a modified balance in VR. By 'modified', we mean a version of the balance model that handles objects with 'negative' weight, while maintaining the affordances of the classical balance model. Negative weights apply an inverted gravitational force to the balance pan, while positive weights apply the standard gravitational force. The virtual environment must enable the user to experiment with and solve linear equations by offering transformational tools. We intend to centre the interaction with the modified balance model on what we call an *invert mechanic*. This invert mechanic enables the user to transform a virtual object to have the additively inverse weight. Positive and negative weight should look and feel the same, but apply opposite gravitational force to the balance pan. Furthermore, the object should be labelled to inform the user whether it has positive or negative weight.

The teaching experience was performed by the authors with ten Danish lower secondary school students using the modified balance in a virtual environment. We have named this environment *Equation Lab*.

Theoretical Constructs

In this section, we present theoretical findings from related work in students' difficulties with linear equation solving, the affordances and limitations of the classical balance model, virtual manipulatives and the role of VR in this context. These theoretical considerations form the design principles for the VR application Equation Lab. Before we present the findings and considerations going into the design of Equation Lab, we introduce the theory of conceptual fields.

The Theory of Conceptual Fields and the Notion of Scheme

Through the work of Gérard Vergnaud, we are familiarised with the theory of conceptual fields. This theory concerns a structure that encapsulates the notion of a scheme and its role in the activity. A conceptual field essentially consists of a set of situations or classes of situations that pairs with a set of concepts or categories of concepts. The actual conceptual field is the individual's interplay between these sets (Vergnaud, 2009).

A conceptual field can also be thought of as a developmental theory:

As a part of the theory of conceptual fields, we became familiarised with how the scheme as a concept works as an organiser of action or activity when engaged in a situation or a class of situations:

Such a situation or class of situations could be exemplified as working with algebraic expressions or engaging in solving linear equations. If we accept that schemes are organisers of the activity of an individual, we can create assumptions about students' schemes by observing their actions and activity in desired situations. Ahl and Helenius (2018) argue that this is the reason why schemes are both didactically and analytically more interesting than the idea of conceptual understanding. Essential to the schemes from Vergnaud's perspective are the operational invariants (the epistemic aspect of schemes), consisting of concepts-in-action and theorems-in-action.

A concept-in-action is an object, a predicate or a category that is held to be relevant by an individual (Vergnaud, 1988).

Vergnaud (2009) defines a scheme as having four aspects (often referred to as 'components'). The first aspect is the intentional aspect, which involves establishing one or several goals and/or anticipations. Second is the generative aspect that involves the emergence of rules that generate the activity or the sequences of actions, including information gathering. The third is the epistemic aspect regarding the operational invariants, referred to as the *concepts-in-action* and *theorems-in-action*. The function is to choose the relevant information and infer from its goals and rules. Last is the computational aspect that involves possibilities of inference. He emphasises that it is essential to understand that thinking is comprised of intense activity of computation, in apparently simple situations and, even more, in new situations.

In every mathematical action, we choose certain objects, predicates or categories of such that are held to be relevant in the current situation or setting. To an individual, a theorem-in-action is a proposition held to be true. When we engage in mathematical situations, we hold 'theorems' to be true or false about the objects that are relevant (concepts-in-action) in the current situation. Vergnaud argues that there is a dialectic connection between theorems and concepts. This comes from the fact that more advanced mathematical concepts emerge from theorems and vice versa. We emphasise that Vergnaud's (1988) interpretation of the representation is similar to what others call a conception or a concept image or invoked concept image (Tall & Vinner, 1981).

Concepts-in-action hold no truth value, just relevance to the situation. Theorems-in-action are in their nature true or false. These propositions provide the concepts-in-action with the possibility of inferences to occur. The *rules of action* are not to be confused with the theorems-in-action. Rules of action function as the generative aspect of the scheme. Their purpose is to be appropriate and efficient, but they rely implicit on theorems-in-action (Vergnaud, 1997). Vergnaud (2009) emphasises that schemes are efficient organisers of activity. If they also become effective, the particular scheme could be considered an algorithm.

Students' Difficulties with Linear Equation Solving

In arithmetic, students view the equals sign as a 'do something' signal and not as a proper equivalence relation (Kieran, 1981). Researchers agree that the difference between an arithmetic thought process and an algebraic one is the comprehension of the equals sign (Kieran, 2007). Matthews and colleagues (2012) suggest the following levels of understanding of the equals sign (see Table 1), where a relational understanding of the equals sign is to be considered as an algebraic understanding.

The levels presented suggest that a student should, at a minimum, have a 'flexible operational' conception of the equals sign to be able to solve, manipulate or possibly even understand algebraic equations (Vlassis, 2002).An algebraic or relational interpretation of the equals sign is therefore essential to understand what an equation is and to engage in equation solving, especially when encountering more advanced equations with operations on both sides of the equals sign, such as $3x + 4 = 6x - 8$ or even $14 = 2 + 3x$ (Rhine et al., 2018).

Table 1 Construct map for knowledge of the equals sign as an indicator of mathematical equality. (adapted from (Matthews et al., 2012, p. 224)

Level 4: *Comparative relational*	Successfully solve and evaluate equations by comparing the expressions on the two sides of the equals sign, including using compensatory strategies and recognising transformations to maintain equality. Consistently generate a relational interpretation of the equals sign	Equations that can be most efficiently be solved by applying simplifying transformations: For example, without adding 67 + 86, can you tell if the number sentence '67 + 86 = 68 + 85' is true or false?
Level 3: *Basic relational*	Successfully solve, evaluate and encode equation structures with operations on both sides of the equals sign. Recognise relational definition of the equals sign as correct	Operations on both sides: $a + b = c + d$ $a + b - c = d + e$
Level 2: *Flexible operational*	Successfully solve, evaluate and encode atypical equation structures that remain compatible with an operational view of the equals sign	Operations on the right: $c = a + b$ No operations: $a = a$
Level 1: *Rigid operational*	Only successful with equations with an operations-equals answer structure, including solving, evaluating and encoding equations with this structure. Define the equals sign operationally	Operations on the left: $a + b = c$ (including when a blank is before the equals sign)

When students engage in solving equations, different strategies come into play. Linsell (2009) suggests that, instead of trying to determine how difficult a given equation is, it is more useful to look at the strategies that the students can use. Furthermore, these strategies do not only describe alternative approaches to solving linear equations, but they also represent the stages of conceptual development. This indicates that teaching equation solving strategies and the rationale behind them is beneficial to students' understanding of the concept of linear equations. Linsell proposes a hierarchy of equation solving strategies as an indicator for the ability of an equation solver. The more equation solving strategies a student can apply, the better that student's understanding of the concept of equations is.

The strategies presented in Table 2 are more or less self-explanatory, except for the 'cover-up' strategy. This strategy or technique is based on the idea of working backwards and substituting. The individual covers the term (or expression, in some cases) containing

Table 2 Classification of strategies used for coding (Linsell, 2009, p. 333), based on Kieran (1992)

1	Unable to answer question
2	Known basic facts
3	Counting techniques
4	Inverse operation
5	Guess and check
6	Cover up
7	Working backwards then guess and check
8	Working backwards then known fact
9	Working backwards
10	Transformations/equation as object

the unknown and then solves the equation using the 'cover' as a new unknown. Next, the student solves for the initial unknown by setting the solution found for the cover equal to the cover. For example, one could cover the term $6x$ in the equation, $6x + 5 = 23$, to get $\square + 5 = 23$, realising that $6x = 18$, before solving for x in this new equation.

Teachers occasionally try to help students learn the working-backwards strategy by introducing phrases like 'change side – change sign' (Rhine et al., 2018). However, this strategy can lead to students making errors or getting a faulty conception of the equals sign. Herscovics and Linchevski (1994) found that students tend to detach the minus sign preceding a number. This way, the students' schemes do not handle numbers correctly when doing transformations, such as working backwards or 'change side – change sign' of an equation, leading to them inverting subtraction with subtraction. Equation solving strategies can be considered rules of action when related to the theory of conceptual fields. These rules or strategies might, in many cases, not be sufficient to solve an equation, especially in new or unfamiliar situations.

Vlassis (2002) identifies equations including negatives as especially difficult by focusing on two major difficulties: the detachment of the minus sign (Herscovics & Linchevski, 1994) and the inability to isolate x when it is preceded by a negative coefficient. Vlassis also observes that the didactic cut (Filloy & Rojano, 1989) does not depend on the structure of the equation, meaning whether it is arithmetic ($ax + b = c$) or non-arithmetic ($ax + b = cx + d$). Rather, it depends on whether the equation has been made abstract by the presence of negatives:

> The negatives place the equation ('arithmetical' or 'non-arithmetical') on an abstract level. It is no longer possible to refer back to a concretemodel or to arithmetic. The "didactical cut" does not seem to depend upon the structure of the equation (unknown on both sides of the equation), but upon the degree to which the equation has been made abstract by the negatives. Arithmetical equations with negativestherefore also represent an obstacle for those students who are unable to give them a concrete meaning. (Vlassis, 2002, p. 350)

Vlassis presents a classification of linear equations in four categories based on arithmetic/non-arithmetic distinctions, building on Filloy and Rojano (1989). The presence of negatives adds a level of abstraction to an equation, regardless of whether it is arithmetic or non-arithmetic. Abstract equations no longer refer to a concrete model or to arithmetic. However, solving equations that are detached from a model implies that other obstacles have to be overcome (unknown, negative numbers, etc.) for which the balance model is not sufficient (Table 3).

Table 3 Level of interpretation of negative numbers based on a table from Gallardo (2002, p. 179)

1	Subtrahend, where the notion of number is subordinated to the magnitude (for example, in $a - b$, a is always greater than b where a and b are natural numbers)
2	Relative or directed number, where the idea of opposite quantities in relation to a quality arises in the discrete domain and the idea of symmetry appears in the continuous domain
3	Isolated number that of the result of an operation or as the solution to a problem or equation
4	Formal negative number, a mathematical notion of a negative number, within an enlarged concept of number embracing both positive and negative numbers (today's integers)

When relating negative numbers or negativity in general to the notion of schemes, we must consider situations. Recognising the role and the relevance of the minus sign in different situations, possibly involving linear equations, can prove extremely difficult.

The Balance Model

Traditionally, the balance model has often been used for introducing and teaching linear equations. According to Otten and colleagues (2019), the balance model is used in teaching equations due to three rationales: *equality concept*, *physical experiences* and *learning through models and representations*. The balance model does an excellent job of representing the equals sign. Using it is seen to enhance the understanding of the concept of equality in general. When the balance is levelled, the two sides represent equal values and are thereby interchangeable. The usefulness of the model in demonstrating the idea of the scales facilitates the use of the rule or strategy of elimination of like terms.

Regarding the second rationale, they state that learning about equality through the physical experience of using the balance model is beneficial, giving learners a greater understanding of the concept of linear equations. Studies underline the importance of movement and gestures when working with the balance to develop mental models of mathematical ideas. By offering students experience with manipulation of balance, equality can be recognised, defined, created and maintained Table 4.

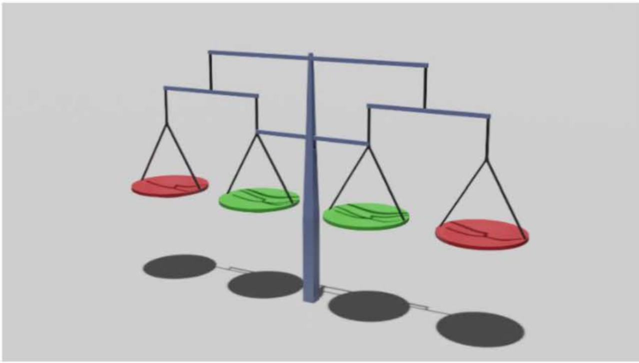
Suh and Moyer-Packenham (2007) emphasise that using manipulable concrete objects can help provide meaning by linking procedural and conceptual knowledge of algebraic equations. However, caution is necessary when using such manipulatives for teaching formal equation solving, because not all students automatically connect their actions on manipulatives with their manipulations of abstract symbols. The real-time feedback, which some models provide, allows students to verify the results of their manipulations and their reasoning processes to construct knowledge. Learning through the use of models and representations is beneficial, because the learner can use the representation of the model to make sense of the abstract algebraic object (Otten et al., 2019).

The traditional balance has no operational signs and prefixes present when looking solely at the visual representation. Therefore, we find it necessary to address the implicit mathematical mechanisms taking place when working with the balance model. When more than one element is present on a balance pan, those elements' values (weights) are naturally added. Additionally, one can then also describe how placing an element on both pans of the balance can symbolise subtraction, if we consider the balanced state of the balance model as representing zero (from a fluctuation perspective). This is not to say that it is using the balance model as a calculator for subtraction, but is merely reflections on what implicit or hidden mathematical procedures to which the balance model provides access. Vlassis (2002) emphasises that the introduction of negatives (both negative terms and negative coefficients) is what detaches linear equations from concrete models such as the balance model.

Pirie and Martin (1997) mention that subtracting a negative number to cancel it out is a common error when learning through the balance model. A further specific difficulty is that a solution needs to be seen as a weight and not just a number. This results in difficulties when negative solutions are encountered. The traditional balance model does not provide the user with the ability to display any linear equation involving subtraction without the terms being semantically different (e.g. helium balloons instead of weights). However, the balance can, in many cases, display an equivalent equation. Consider the equation $8 - x = 5$. This equation can not be displayed in its current form on the balance using the concepts of numbers as weights. The equivalent equation is $8 = 5 + x$. We find it important to emphasise the difference between the minus sign as an operation and the minus sign as a prefix.

On the 4-pan balance, one can represent an equation involving a subtraction (see Fig. 1). By placing the weight on the red (outer) pans, the balance reacts as if that weight is being subtracted from the weight on the green (inner) pan on the same side. In the case of $8 - x = 5$, one could place 8 and x on the left side with 8 on the green (inner) pan and x on the red (outer) and place 5 on the right ride on the green (inner) pan.

Fig. 1 4-pan balance that enables subtraction



Design

This section will present the design of the VR application and the associated didactic considerations. These didactical considerations are mainly described in terms of how the schemes of the learners are engaged and based on the affordances mentioned. The VR application is designed as a virtual manipulative environment (VME) and consists of a balance with two pans and a beam, a linear equation represented by virtual manipulatives (Moyer-Packenham & Bolyard, 2016) distinguished as mathematical terms and an area for the formal transformation of the terms (see Fig. 2).We refer to an interactive feature of the VME that a user can control as a mechanic. The user interacts with the VME through a hand-held controller represented in VR by a virtual hand.

The balance depicts mathematical equality or inequality through two visual cues: the deflection of the balance pan and the colour of the balance beam. In other words,

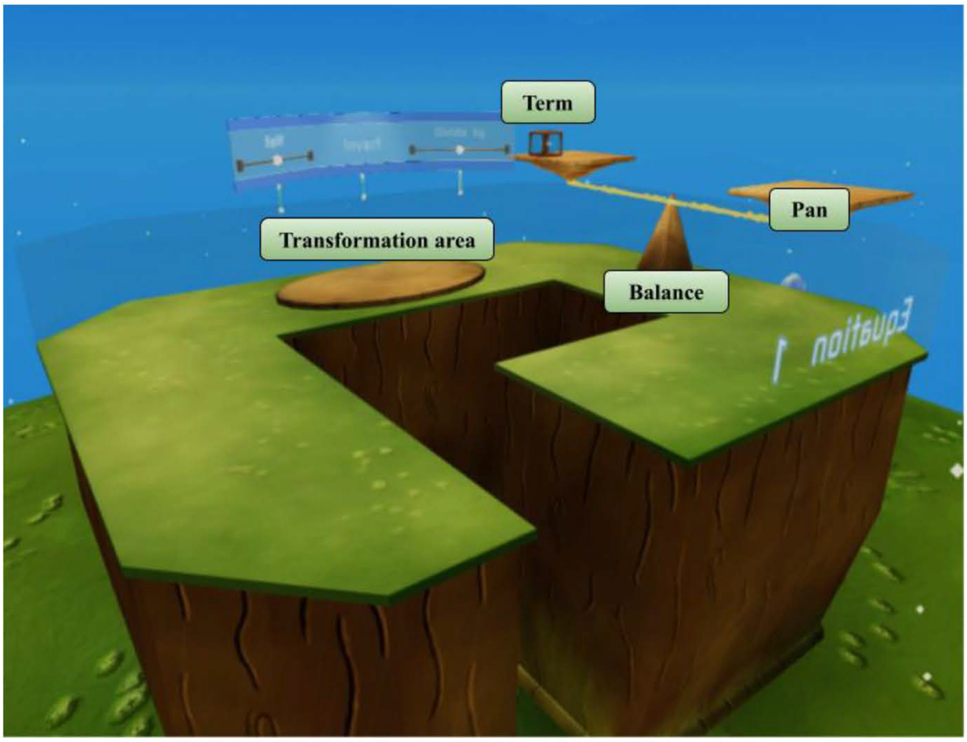


Fig. 2 Overview of VME

the two states of equilibrium and disequilibrium are represented by the gradual deflection from 'weight sums' and the colour coding of the balance beam (i.e. green for equilibrium and orange for disequilibrium). While the beam only illustrates the two states, the deflection can be used for further accuracy in making inferences about experimentations and the outcome of transformations.

Virtual Manipulatives of Equation Lab

In this sub-section, we describe the design ofvirtual manipulatives representing terms of the equations and how we allow for transformations (see the example range in Fig. 3). Individual virtual manipulatives represent terms, which can be moved from the balance andtransformed in a designated area (the functionality of the transformation is described in a later sub-section). Each virtual manipulative contains a dynamic text element indicating the mathematical value of the term (e.g. 4, –6 or $7x$). This also means that a term can contain one or several multiples of the unknown. If a term contains an unknown, the term will act correspondingly to the total value of the term when the coefficient is multiplied with the value of the unknown in the given situation. The value of the unknown is hidden to the user, while the balance reacts according to the term as if the value were known. We designed this behaviour to allow for applying the same transformations on

Fig. 3 An example of terms in Equations Lab

all virtual manipulatives to create the task of solving an equation in the virtual environment.

Representation of an Equation

The terms of a linear equation are represented by virtual manipulatives resembling a framed box with the value of the term displayed inside as a text element (see Fig. 3). The appearance of the terms was designed to resemble each other regardless of being a known number or a multiple of the unknown. An equation is manipulated simply by moving or transforming the terms of which the equation consists. A term is manipulated (transformed) by moving it from the balance pan to the transformation area, where a menu with three formal transformation options appears (see Fig. 4). We wanted to avoid representing negative terms by something semantically different, like a floating balloon, because we felt it would potentially not signify the ability to be placed in the transformation area or the ability to be handled in the same way a box containing a seemingly positive value would.

In addition, we noticed the interesting design difficulty in representing terms with multiples of the unknown in the equation. If a positive element should be represented semantically different from a negative element, then the element $-6x$ would have to be represented differently if the unknown were a positive number rather than a negative number. While we shall discuss this later, we initially chose to represent terms with a negative value in the same way as ones with a positive value, because we believed that this design choice would increase the users'

Fig. 4 Procedure for transforming an equation term

opportunities for experimentation leaning into the affordances of learning with the balance model (Otten et al., 2019).

The intention of the design is for the users to engage with the situation of linear equation solving, which involves several steps of formal transformation by manipulating the terms in the designated transformation area. We intend for the students' equation solving schemes to be strengthened by creating equation solving situations in which inferences about the role and function of negative values can be made.

Dynamic Balance

The overall aim of this study was to modify the balance model to afford experimentation with negative quantities. For this reason, we have designed our virtual balance to react differently from the way a regular balance would. The terms represent both negative and positive quantities and thereby also both negative and positive weights, but are encapsulated by the same 'box model'. Therefore, the virtual balance is designed to react 'correctly' to the application of negative weight, by reacting as if the box were applying reversed gravitational force to the balance pan pulling it upwards.

When applying a positive weight, the balance reacts 'normally', in the sense that the gravitational force of the box is applied to the pan pushing the pan downwards. The balance is intentionally implemented as a dynamic model to afford real-time feedback (Otten et al., 2019). The dynamic model of the balance deflects in real-time based on the presence of terms of the equation on each pan. The freedom of placing and removing terms allows students to experiment with and experience the effect of the modified balance and how it reacts when moving terms from one side to the other. For example, a student's schemes can adapt to the fact that the deflection effect of the term 5 on one side has the same as the deflection effect of the term -5 on the opposite side.

As already mentioned, no operational sign is present on the balance. As an example, the equation $7x - 3 = 18$ would be represented as the terms $7x$, -3 and 18, with $7x$ and -3 on one pan and 18 on the other. When working with the balance, objects are considered as representing a weight. This intuitively suggests that the weights of all the objects present on one pan should be seen as one collected weight and therefore should be added together. This adds some level of complexity to the balance model.

Transformation of Terms

When performing transformations on terms, the user can choose from the following actions: *split*, *invert* and *divide*. These three transformation possibilities are explained in detail in the following. In the case of manipulating multiple terms, an additional option of *add* is given (see Fig. 5).

If a single term is placed into the transformation area, the user can choose to split the terms into two terms that sum up the original term. In the case where the user

Fig. 5 Actions in the transformation area

splits the term 5, this could, for example, choose and yield the term 2 and the term 3. In the case where the user splits the terms $7x$, the user could choose and yield x and $6x$. The user manipulates a slider to set the desired split values before executing the command by pulling the corresponding handle (see Fig. 6).

Fig. 6 Adjusting the value and executing the action

This transformation option has been implemented for the user to create a situation where like terms are present on the balance model. If the user is faced with the equation $7x = 3x + 16$, the user could then split the term $7x$ into the terms $3x$ and $4x$. Then the user could infer that the term $3x$ is present on both sides of the equation (i.e. both balance pans) when the balance is levelled. This could result in the user removing both terms containing $3x$ to be left with the equivalent equation $4x = 16$ as mentioned as an affordance of using the balance model.

The user can also choose additively to invert the value of a term (see Fig. 7). This effectively enables the user to manipulate or alter the appearance of entire equations,

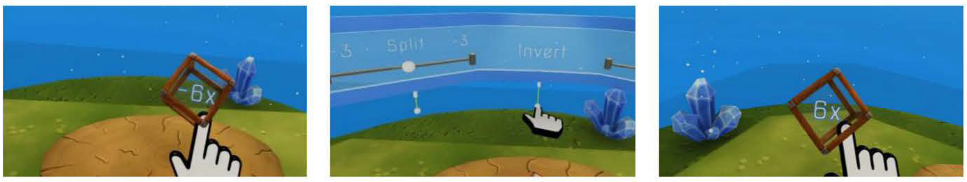


Fig. 7 The action of the invert mechanic

by inverting the sign of every term on both sides, in order to yield an equation that resembles one from a more familiar situation. This could, for example, be done by transforming the equation $-6x = 24$, which causes students difficulties, into the equation $6x = -24$. Alternatively, the invert mechanic could be used to enable the student to move a term from one side of the balance model to the other and still maintain balance. If a term is inverted before being placed on the opposite balance pan, then the user is enabled to collect similar terms.

In the design process, we carefully chose to implement the invert mechanic because we wanted to avoid implementing multiplication as a transformational option. Multiplication by –1 on both sides of an equation could arguably be a more well-known procedure for students capable of performing formal transformations on equations where the last term containing an unknown has a negative coefficient. Because we wanted to avoid the user multiplying terms by numbers other than –1, we were left with the choice either to multiply by –1 and only –1 or to come up with an alternative. We were quite certain that the students in Danish lower secondary school would be unfamiliar with the word 'invert'. At least in the word's mathematical sense. Because the terms represent weights when associated with the balance model, we decided to implement the ability to invert instead of the ability to multiply (or divide) by –1. Through the study and discussions presented in later sections, we reflect on this and the other design choices.

The last single term transformation is the ability to divide a term by a positive true divisor larger than 1. This is most useful when the user is left with equations like $ax = b$. In this situation, the user can choose to divide the terms ax and b by the absolute value of the number a to achieve the terms x or $-x$ in the case where the number a is a negative number. Indeed, this transformation could be accomplished by using the split mechanic. However, this would be done by splitting the term ax into the terms $(a-1)x$ and x, in the case where the number a is a positive number. However, if the user had only the split mechanic to perform the same transformation on the term b, the user should have to perform the calculation in the head before choosing the right amounts into which to split the term b.

We find this procedure of using the split mechanic as a tool for the division to be an unhealthy way to perform division. For the sake of establishing a mechanic that worked similar to pen-and-paper situations, and because we believe it is easier to remember a divisor, we have implemented the divide mechanic. The user is left with terms summing up to the original term when performing the division. For example, the user could choose to attempt to perform division on the term $6x$. In this case, the values the slider would be able to choose would be $\{2,3,6\}$. If the number 2 was chosen as the divisor, the user would be left with the terms $3x$ and $3x$. Similarly, if the user chose the number 6 as the divisor, she would be left with six terms all with the value of x. This is to enable the user to perform a sub-optimal division for the sake of experimentation.

The add mechanic serves multiple didactical purposes, as well as being the only mechanic enabled when multiple terms are present in the transformation area at the same time. When multiple terms are placed in the transformation area, the user interface changes to only show the option of addition. The add mechanic lets the user collect like terms, e.g. 4 and 6 and merge them into 10. Likewise, this can be done to terms containing an unknown, e.g. $3x$ and $-7x$ yielding $-4x$. The main purpose for designing an add mechanic was to help the user simplify the equations on the balance by merging like terms before placing them back on the balance. However, the add mechanic also lets the user undo undesired splits or divisions during experimentations with the equations. The add mechanic is the only mechanic out of the four possible transformation mechanics that let the user destroy content from the equations. In some cases, the user could add two terms that are of the additive inverse values, e.g. 4 and −4. Adding additive inverse terms results in irreversible experimentation. If the user would attempt to perform addition on non-like terms, then the user interface would show a warning message saying 'incompatible' and proceed not to carry out the addition.

The ability to transform terms within the virtual environment expands the possibilities for how equations can be represented compared with the traditional balance model. Without the ability to divide or split terms, we would only be able to represent the equation $3x = 9$ as $x + x + x = 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1$ (three x's on one pan and nine 1's on the other) to enable the user to verify that x is equal to $1 + 1 + 1$ or 3. The balance in our virtual environment, however, does not afford subtraction as an operation. The classical balance model provides the intuitive feeling of addition when items are placed on the same pan and thereby applies a collected force to the pan identical to an item with the combined weight of the objects placed. As mentioned, we did not choose to implement the 4-pan balance, but rather the classical balance model, since we believed that it would be more readily recognised by students as a tool for displaying differences between weights placed on the balance pans.

We emphasise the importance of addressing subtraction in our virtual environment and with the balance model in general. Since we did not choose to implement a 4-pan version of the balance model, subtraction is not possible to represent on the balance. Negativity is present in the environment as a prefix on a term that signifies the directional force (directed number — Gallardo, 2002); the term will apply to the balance pan. Therefore, subtraction is only representable as an added value of a negative term. This means that, when representing the equation $7x - 5 = 33$, the equation is represented as $7x + (-5) = 33$ on the balance. The intention is for the user to accept that the collection of terms on a pan applies the sum of the force, meaning that, in the case of the above equation, the terms $7x$ and (-5) apply a collected force of 33 downwards (same as the term 33).

A therefore obvious limitation of our design is an issue that persists from the traditional balance, namely that the equation $7x - 5 = 33$ on pen-and-paper is different from the one we can represent within the virtual environment. The idea of a directed number (Gallardo, 2002), represented as a directional force applied to a balance pan, might help users adapt their schemes to how negative numbers can be interpreted and how to work with negativity when solving linear equations. When combined with the invert mechanic, users have an opportunity through experimentation to make sense of difficult equations such as $-6x = 24$ and develop schemes to solve them based on directional force.

An Exploratory Study

In the following sub-sections, we present the methodology and findings from an exploratory study based on a teaching experience using Equation Lab. In the 'Method' sub-section, we explain how data was gathered and how the teaching sequence was set up. After that, we present a sub-section on the procedure of the teaching sequence that was performed and then one presenting data gathered before, during and after the experience. Finally, we discuss and conclude how the gathered data provided responses to the research question.

Method

The participants in the study were from a Danish year 8 class in conjunction with their male teacher. The math teacher of the class selected the ten participants based on gender, performance and experience with playing digital games. Gender and performance level were the most important for us. The part of the study where the students tried Equation Lab took 3 days. During these days, two of the initially chosen students did not want to participate, and so two other students were chosen to fit similar profiles. Fortunately, this proved possible, since all the students had taken the pre-test a month in advance, which consisted of ten items including different kinds of equations. We performed teaching sequences with six boys and four girls using Equation Lab. The data collection consisted of three parts (see Fig. 9): pre-test, teaching first with VR, then using pen-and-paper, and, finally, interviews.

The students were to participate in the teaching sequence with Equation Lab one at a time and with only that one student present. The teaching experience was filmed using mixed-reality capture, which is a technique that allows filming the student in the virtual environment, using a green screen and a virtual camera in combination with a real camera (see an example in Fig. 8). The authors acted as teachers (research teachers from now on) during the teaching sequence with Equation Lab. Afterwards, the students attempted to solve the same equations as they were working with in VR, but using pen-and-paper and a classic symbolic representation of the equations and still with some minor guidance from the research teacher. While

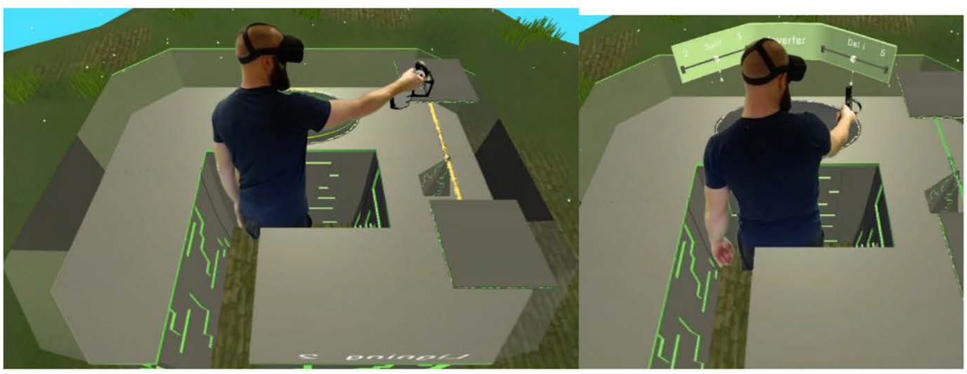


Fig. 8 Example of mixed-reality capture from Equation Lab (early graphics)

solving these equations using pen-and-paper, the research teachers would reflect on the process of solving the equations in VR and how we, as researchers, could hypothesise about the students' equation solving schemes involving pen-and-paper compared with those of the VR environment.

After these teaching situations, the math teacher whom the participants normally worked with agreed to focus on linear equations for the following 2-week period. This sequence was not done using Equation Lab and, without the presence of the authors, the math teacher would normally teach linear equations in this manner. After these 2 weeks, we interviewed the mathematics teacher on his observations on students' behaviour following the teaching experience. The teacher pointed to one student, and, therefore, we conducted an additional interview after talking to the teacher (Fig. 9). We present the results of this interview in the later sub-section called 'The Case of Albert'.

Video material obtained during the exposure to the VR experience was used for transcription and analysis of the students' equation solving schemes in regard to the modified balance and exercises on pen-and-paper with a special focus on negative numbers. The video material documented both the VR experience and the pen-and-paper exercises. The teaching sequence was largely a guided experience and, therefore, we focus on the pen-and-paper exercise, in which students were allowed to reflect on the concepts from the VR experience while working with equations on pen-and-paper.

Students were given a brief introduction to the experience and the user controls within it. After the introduction, the students were fitted up with the VR equipment. As described in the 'Design' section, we were able to present the user with six scenarios *centred* on equation solving. Each student went through each scenario before performing a pen-and-paper exercise solving the same equations as in Equation Lab. In the following, we describe the six scenarios.

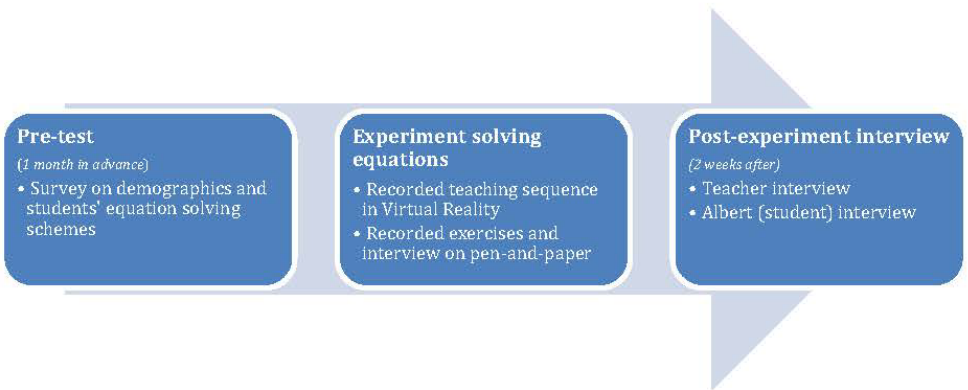


Fig. 9 Diagram of collected data

The Six Scenarios in Equation Lab

The following sub-section presents the rationale behind the included scenarios (tasks) of Equation Lab. The user is presented with the following equations (inequalities as introductory experience) as six scenarios: $-4 < 0$, $5x < 0$, $3x = 9$, $7x + 5 = 33$, $-6x = 24$, $7x - 3 = 13x + 15$.

Introductory Experience

In the first scenario, the user was only presented with one object, namely – 4 placed on the left pan. This does, of course, ultimately result in the representation of the inequality $-4 < 0$. The user would face a balance that was not levelled. The goal of this task, and the next one, was not for an equation or an inequality to be solved. The goal was merely to enable the students to familiarise themselves with the environment and the properties of the terms and the virtual balance.

The second scenario presented the user with the object $5x$ sitting on the left pan, resulting in the inequality $5x < 0$. This scenario extends the first by providing the user with an opportunity to familiarise themselves with how the environment handles unknowns and the possibilities unknowns present. The unknown was set to –2, which resulted in the object ultimately behaving like –10. This was chosen because the user could draw on the experience from the previous scenario, remembering how the balance reacted to negative weights (terms). The user would also be able, based on the first introductory task, to make assumptions on whether the unknown should be a positive number or a negative number.

In both introductory tasks, we wanted the user to experience the dynamic balance and the workflow of using the transformation area. The user would get the opportunity to work on a term containing an unknown and a term with no unknown present.

Equation Solving Tasks

In the third scenario, the user was presented with an equation, namely $3x = 9$. The term containing $3x$ was sitting on the left pan and the term 9 on the right. The balance let the user know that it was levelled. This equation was chosen as a supposable easy task that could give the user a taste of what was coming. The research teacher had the opportunity to guide the user to a solution, utilising the affordances of the balance model as an alternative to guessing that the solution was 3.

The fourth scenario presented the user with the equation $7x + 5 = 33$. The terms $7x$ and 5 were sitting on the left pan and 33 on the right. This equation was meant to be a sensible step further than the equation before. This scenario let the user to experience a situation where removing the term 5 on the left pan lead to an equation like the one in scenario three. The user could either split 33 to remove 5 from both sides (using the common affordance of the balance model) or invert the term 5 and place it on the right pan to experience the power of 'change side – change sign' in a dynamic virtual setting. Of course, the user would have to add the terms 33 and -5 to achieve 28.

The fifth scenario consisted of the equation $-6x = 24$. This equation was chosen because, traditionally, equations like this can neither be solved nor represented using the classical balance model (Vlassis, 2002). However, as a part of the overall aim, we wanted to explore what happened if the students were to work with equations made abstract by negative coefficients in Equation Lab. Here, the user could take two routes: either divide both sides by 6, to end up with $-x = 4$ or invert each side to end up with $6x = -24$. Both routes involved inverting each side and dividing each side by 6, but with the order of transformations reversed. The invert mechanic becomes essential to solving this type of equation using Equation Lab, because we only allow division by positive numbers and the intention was to observe the students' schemes in situations solving equivalent equations, namely $-6x = 24$ and $6x = -24$.

In the last scenario, the user was faced with the equation $7x - 3 = 13x + 15$ (Bodin, 1993). This is, in many ways, a classic non-arithmetical equation (Filloy & Rojano, 1989) and also an equation 'detached from models' (Vlassis, 2002), because the solution is a negative number (–3). This equation in VR does distinguish itself significantly from the equation in pen-and-paper form because of the missing operational signs. For these reasons, it was included in this teaching experience. Additionally, the equation has the term -3 on the left side (pan) which might cause students difficulties related to the 'detachment of the minus sign' (Herscovics & Linchevski, 1994).

Furthermore, to avoid negative coefficients, it could be sensible for the user to collect the unknowns on the right side (pan), which could seem illogical to some students who would prefer to end up with an arithmetical equation like $ax = b$, with the unknowns collected on the left-hand side. If the students collected the unknowns on the left side, they would end up in a situation similar to the fifth scenario ($-6x = 18$). The students could also end up with $-18 = 6x$ if the strategy of collecting like terms was done the opposite way. In any case, we did not expect the students to attempt to guess the solution to this equation.

Results

In this section, we present the results of the exploratory study and commentary related to the research question. We start by presenting a summary of the results from the pre-test before going into responses and reactions the students gave when reflecting on the VR experience while solving the same equations on pen-and-paper during the teaching experience. Hereafter, we present a summary of the results regarding the participants as a group and focus on results from an interview with Albert, the student highlighted by the math teacher of the class, 2 weeks after the teaching experience.

Pre-test

In this sub-section, we give an overview of how the participant students handled the tasks from the pre-test. Its purpose was to get a feeling for how the students' schemes would engage with, solve or strategize about different tasks related to the concept of linear equations. The following tasks were part of the pre-test that was sent to the math teacher of the class 1 month before the teaching experience with Equation Lab. The students solved the tasks individually and without help from the teacher.

1. What number goes in the empty space $8 + 4 = \_\_\_ + 7$?
2. What is the value of x, if $3x = 15$?
3. What is the value of x, if $14 - x = 6$?
4. What is the value of x, if $8x + 5 = 21$?
5. What is the value of x, if $3x - 4 = 5x - 12$?
6. What is the value of x, if $12 - x = 15$?
7. In class, you solve the equation $5x - 3 = 23$. Three of your classmates show you their attempt at the first step towards a solution. Asta shows you $5x = 20$, Tanya shows you $2x = 20$ and Carl shows you $2x = 23$. Do you agree with one of them or none of them?
8. What is the value of x, if $7 = 3 - x$?
9. The bags contain the number of marbles corresponding to the number printed on them. What should be the number on the bag labelled x?
10. The bags contain the number of marbles corresponding to the number printed on them. What should be the number on the bag labelled x?

Table 4 is an overview of how the student participants answered the tasks listed above. Student H, unfortunately, was unable to take the pre-test. In the table, we highlight both the answer and strategy used, if available.

Table 4 Scores from the pre-test (light green signifies a correct answer, while dark red signifies an incorrect answer).

Task	Student A	Student B	Student C	Student D	Student E	Student F	Student G	StudentI	StudentJ
1	*5*	*5*	*5*	*12*	*5*	*12*	*5*	*12*	*12*
2	*Counting strategy*	*Counting strategy*	*Guessing/ Counting strategy*	*Counting strategy*	*Formal transformation*	*Guessing / Counting strategy*	*Guessing/ Counting strategy*	*Guessing / Counting strategy*	*Counting strategy*
3	*Counting strategy*	*Counting strategy*	*Guessing strategy*	*Counting strategy*	*Formal transformation*	*Guessing strategy*	*Guessing strategy*	*Guessing strategy*	*Counting strategy*
4	*Guessing strategy*	*Guessing strategy*	*Guessing strategy*	*Guessing strategy*	*Formal transformation*	*Guessing strategy*	*Guessing strategy*	*Guessing strategy*	*Guessing strategy*
5	*Guessing strategy*	*Guessing strategy*	*Guessing strategy*	*Incorrect*	*Guessing strategy*	*Incorrect*	*Does not accept*	*Does not accept*	*Does not accept*
6	*Correct*	*Does not accept*	*Correct*	*Does not accept*	*Reverses order*	*Reverses order*	*Reverses order*	*Reverses order*	*Reverses order*
7	*Asta*	*Correct*	*Correct*	*Correct*	*Correct*	*Carl*	*Correct*	*Correct*	*Asta*
8	*Guessing strategy*	*Does not accept*	*Correct*	*Correct*	*Reverses order*	*Reverses order*	*Does not accept*	*Reverses order*	*Reverses order*
9	*Correct*	*Correct*	*Correct*	*Correct*	*Correct*	*Correct*	*Correct*	*Correct*	*Correct*
10	*Correct*	*Correct*	*Correct*	*Correct*	*Correct*	*Correct*	*Correct*	*Correct*	*Correct*

To sum up the results from the pre-test, we went into the teaching experience with Equation Lab, anticipating that the students would not be confident nor have particular efficient schemes in situations working with negative numbers in relation to equations. The tasks that involve negative solutions (tasks 6 and 8) are arguably quite difficult. We wanted to investigate if the students were willing to subtract negative numbers in task 6. In task 8, we took it a step further, by having the operation on the right-hand side of the equals sign. This does require the students to be a little more flexible in their understanding of the equals sign (Kieran, 1981; Matthews et al., 2012).

Furthermore, we saw that guessing strategies (trial and error) were widely used for the items we presented in the pre-test. We noticed that the students did accept the balance as a metaphor for equality. All the students were able to work out how many marbles should be in the bags with an x on them. We also noticed that four out of nine students were not able to see that the number 5 goes in the empty space in task 1 (Falkner et al., 1999).

Summary of Participants' Reflections of Solving Equations in VR While Solving Equations on Pen-and-Paper

In the post-VR, pen-and-paper exercise, students were asked what they could use from the VR experience and take it with them to the classroom regarding equation solving in its classical form. In the following, we present comments and realisations made by the students during the pen-and-paper interview after the VR experience. In this part of the teaching experience, the students reflected on the VR experience while solving the same equations on pen-and-paper. We outline tendencies between participants concerning realisations of mathematical equality and equation solving strategies and their subjective impression of what they gained from trying out the VR experience.

As expected, we could infer from the pre-test that most of the students, except for the students A and C, did not work comfortably with negative numbers. Almost half the participating students did not have a sufficient understanding of the equals sign to give a correct answer to task 1 ($8 + 4 = \_\_\_ + 7$). As predicted, none of the student participants was familiar with the word *invert* in its mathematical sense.

When working with the students, we wanted to experience the VR application (Equation Lab) as a tool for teaching linear equations. In particular, it was teaching linear equations made abstract by negatives (Vlassis, 2002). We felt it was easy to get the students to work with the balance and accept the metaphor of terms as weights pushing or pulling on the balance pans. The invert mechanic naturally became a centrepiece in teaching experience. Most of the students felt that the invert mechanic helped them make sense of the equation $-6x = 24$. In our experience, the students would not have been able to make sense of this equation before the teaching experience. Several of the students did refer to motions and concepts from the VR experience when working with the equations on pen-and-paper.

When explicitly questioned about the added value of VR, several students reported satisfaction with the system's transformation options and ability to help with arithmetic calculations, potentially allowing focus on equation solving strategies. In addition, students reported that the VR experience provided them with a mental image of the balance model that they could access and use to solve equations in the classroom. Moreover, students reported an increased self-confidence when working with equations in VR compared with using pen-and-paper.

Teacher Interview

In agreement with the mathematics teacher of the class, we proposed a 2-week teaching programme on equation solving not influenced by the authors. Following the 2-week teaching programme, we interviewed the teacher of the class on the experience gained after students had been exposed to our teaching sequence in VR. The teacher reported on the general competence level of the class and pointed out a specific case of interesting behaviour.

He reported that, in general, the students in his class were at an expected competence level. However, he described a particular student's strategy in equation solving as interesting and convoluted, and described how it was clear that the experience in VR had affected this student's strategy (schemes) for solving equations.

> Albert's approach was rather interesting. He would split every element of the equation up [like in Equation Lab] before gathering each element again. As a spectator, it was confusing to follow, but it seemed to make sense in his own mind, since at the end of the equation-solving task, he would reach a correct solution.

The strategy was divergent from an algorithmic approach that the teacher had introduced to the rest of the class. However, it proved to produce correct results in exercises that the student was presented with. The teacher described the strategy as confusing, as it was his impression that the student was using a mental image of the environment and tools of Equation Lab. The student used similar actions, such as splitting and moving terms. He also explicitly used the terminology from the VR experience by referring to a 'change side – change side' strategy by inverting a term.

> He used the thoughts and processes from VR, where he was allowed to move and transform elements in a designated area before moving them back to the balance.

The student also showed an emotional response of irritation when asked to use a 'simpler' strategy, as it was the teacher's impression that the work of the student could be confusing to the rest of the class, who did not have access to the same strategies as the particular student.

> He got frustrated when he was asked to use a simpler strategy that I had taught the class, since we were just in the early stages of defining what an equation is.

In the following sub-section, we report the behaviour of the student in question, Albert, the student who showed to gain access to new equation solving schemes from using the VR application.

The Case of Albert

In this sub-section, we highlight the behaviour of Albert, a student who, at first glance, responded very well to the teaching sequence with Equation Lab. In an interview 2 weeks afterwards, the math teacher of the class participating in the study pointed out that Albert (a boy) showed some interesting strategies (we note these as equation solving schemes) when solving equations in class during the 2 weeks of working with linear equations immediately after the teaching experience with VR. In class, Albert had used the invert mechanic when solving more abstract or algebraic equations similar to $-6x = 24$ and $7x - 3 = 13x + 15$.

Initially, we highlight some of the important observations from Albert's pre-test, leading to observations from the teaching sequence and finally additional observations from the interview 2 weeks after the teaching sequence. During the pre-test, Albert did not seem to have any problem with solving less-abstract arithmetic equations with positive solutions. These equations consisted of the following: $3x = 15$, $14 - x = 6$, $x + 5 = 21$, $3x - 4 = 23$. He was able to solve these equations using a guess-and-check or counting strategy (Linsell, 2009). There was no sign that he had considered any other strategies.

The pre-test featured more abstract equations with negative number solutions, namely $3x - 4 = 5x - 12$, $12 - x = 15$, $7 = 3 - x$. Albert was not able to solve these more abstract equations. He did attempt to apply a guess-and-check strategy when solving $3x - 4 = 5x - 12$, but had to give up during the process. When solving $12 - x = 15$, Albert gave $x = 27$ as the solution. We expected some students to provide this answer and he was not the only case. Subtracting the smaller number from the larger is a common mistake when reading an expression like this, done to avoid dealing with negative numbers (Gallardo, 2002). Our initial analysis of Albert's equation solving schemes is therefore limited in terms of rules of actions (strategies) and the computational aspect, regarding inferential power, since he was not able to infer that $12 - 27$ does not equal 15.

As mentioned in the descriptions of the scenarios, the fifth and sixth scenarios of the teaching experience involved solving the following two equations: $-6x = 24$ (Vlassis, 2002) and $7x - 3 = 13x + 15$ (Bodin, 1993). These particular equations fall under the abstract category (Vlassis, 2002) and they were chosen because they are great examples of an arithmetic and a non-arithmetic (Filloy & Rojano, 1989) equation that are turned abstract by the presence of negatives. The negative numbers play several roles in these equations since they both have negative solutions and negative coefficients in the case of $-6x = 24$, present from the outset. As a reference, 36% of first-year upper secondary school students in Denmark are not able to solve the equation $-6x = 24$ (Jankvist & Niss, 2017).When Albert tried the first two actual equations in Equation Lab (scenario three, $3x = 9$ and four, $7x + 5 = 33$), he was immediately able to guess the solutions.

Based on Albert's results from the pre-test, we had expected no less from him. On a side note, this was true for several of the students trying out the experience. When asked why the equation $-6x = 24$ looked especially difficult, Albert answered that the minus sign in front of the number 6 was the reason. He was, in this case, not able to guess or use a counting strategy to get to the solution. He was not conflicted on setting a goal for the solution of the equation, he was however not able to establish relevancy for or a theorem about the minus sign preceding the number 6. We had observed that he was more than able to solve equations such as $ax + b = c$, whare a and b are natural numbers. On the other hand, he was immediately confident that the environment would allow him to experiment towards a solution for this difficult equation. When solving $-6x = 24$ in Equation Lab, Albert made an interesting observation. The following is an extract from the teaching sequence using Equation Lab. Albert was at the stage of the equation where he had divided by 6 on both sides and was left with $-x = 4$:

1	Albert	We can invert. [Moves $-x$ to the workspace and inverts it. Then he grabs the box with x and places it on the left pan.]
2	Researcher	Did the cord turn green? [Refers to the line between the pans that helps the user know if the pans are levelled.]
3	Albert	No, because x is less than four, it [the balance] says
4	Researcher	That was a good observation

Albert was able to utilise the power of the dynamic balance to infer something about the value of the unknown. After this observation, he further observed that it was necessary to invert both sides — as he had done before in Equation Lab — to obtain balance once more. Here, we observe that Albert's experimentation in VR helped him to infer something about the directional force of the unknown.

In the interview 2 weeks after the teaching sequence, the math teacher pointed out that Albert had shown interesting behaviour, taking the experience of Equation Lab to heart when solving equations in class. He had used the invert mechanic when performing transformations on equations similar to $-6x = 24$ and $7x - 3 = 13x + 15$. We wanted to observe for ourselves how Albert had adapted his equation solving scheme from the experience with Equation Lab. During this interview, we also had him solve some additional equations using pen-and-paper. Here, he was still not able to guess the solution to the equation $-5x = 25$, but he was able to solve it using the invert mechanic. Inverting each side before having an additional guess helped him (in fact, students participating in the teaching experience tended to regard the equation $6x = -24$ as much easier than the original, $-6x = 24$).

An important observation, in our opinion, is that Albert's experience with equations such as these still did not let him confidently adapt his scheme to infer that the solution was a negative number. Only when the invert mechanic had helped him effectively to move the minus from the coefficient of the unknown to the number on the right side of the equals sign. Albert was then able to see that, in order to have five times an unknown number result in a negative number, the unknown number itself had to be negative. Looking at the original equation, $-5x = 25$, his guess would have been that 6 was the number that solved the equation. This is maybe not obvious, but this is because $5 \cdot 6 = 30$ and $30 - 5 = 25$. The number entry 5 gets a dual meaning due to the added prefix minus sign.

Albert seemed to be able to apply the version of 'change side – change sign' from Equation Lab, enabled by the invert mechanic. At the interview 2 weeks after the teaching experience, he was not completely successful while solving the equation

$4x - 10 = x + 2$. He ended up with $3x = -8$ after attempting to collect corresponding terms on either side. After explaining to him that 10 should be added and not subtracted on the right side, we asked him why he had chosen to subtract 10. The following is an extract from the interview when Albert was solving the equation $4x - 10 = x + 2$ and explaining why he wanted to subtract 10 from 2 on the right side. He had already collected the unknowns on the left side.

5	Albert	Then I want to write minus ten
6	Researcher	Minus ten, why?
7	Albert	Because I want to move it [10] and because it's positive there [pointing to 10 in the original equation]
8	Researcher	It [10] is considered positive there
9	Albert	Yes, it's positive
10	Researcher	I have to tell you that it is not. Why do think that is?
11	Albert	I don't know
12	Researcher	Because of that minus sign there [pointing to the operational minus preceding 10 in the original equation]
13	Albert	There is a lot of space there
14	Researcher	So, there is too much space between the minus sign and ten
15	Albert	Yes

Albert's reaction was that the minus sign preceding 10 was not close enough to the number 10 for him to consider our proposal. Albert was correct in the sense that 10 is considered a positive number that is subtracted from $4x$. We (he and the researcher) could have agreed that the number was positive and proceeded to talk about how this operational minus sign was different from the minus sign that was a prefix of the terms in VR. It seems that, in this case, it was not the strategy for solving the equation that caused Albert problems, but instead it was the inferences about how the algebraic notation links to the strategy of inverting and moving to the other side or the classical 'change side – change sign' method. This is a notable observation, which should be addressed in future designs and evaluations to avoid hiding important concepts of equation solving.

Discussion

We set out to investigate how a modified balance model in a VR environment could strengthen students' schemes for solving linear equations with a special focus on negativity and negative numbers. In addition, we wanted to explore how the built-in 'invert mechanic' could further assist the students in their development of strategies for solving linear equations involving negativity.

To discuss how the modified balance could support the students' schemes for equation solving and their equation solving strategies (rules of action), based on the teaching experience, some key features were in play. Before the teaching experience, we were curious how the participating students' schemes would adapt to the invert mechanic, mostly due to the unfamiliarity with the concept of inverting. As highlighted in the feedback given in the results from the interview during the pen-and-paper exercises, we consider the invert mechanic as a feature to which the students' schemes were able to adapt.

The students were quick to grasp the usefulness of transforming equations such as $-6x = 24$ into $6x = -24$ — an equation about which they were, to a higher extent, able to make inferences. Through experimentation, Albert was able to adapt his schemes to make inferences about how numbers influenced the modified balance as virtual weights (directed numbers). In fact, several of the participants were able to adapt to the idea that pressing down on one side had the same outcome as pulling up on the opposite side.

Throughout this article, we have mentioned the 'change side – change sign' strategy several times. Results show that the students, especially Albert, were willing and able to apply our new version of the 'change side – change sign' strategy enabled by the invert mechanic in the pen-and-paper exercises. Students explicitly referred to the 'invert' mechanic from the VR experience as something that was a positive and understandable takeaway. The invert mechanic essentially enabled the students to work with and make sense of equations, made abstract by negativity, similar to the equations from scenario 5 ($-6x = 24$) and 6 ($7x - 3 = 13x + 15$).

The invert mechanic presented us with overall observations. The first was the capability to present the students with a situation where they were able to infer that the equation $-6x = 24$ could be legitimately transformed into the equation $6x = 24$. We saw from the interview with Albert that equations such as $-6x = 24$ ($-5x = 25$ in the interview) are subject to interpretations that are not necessarily understood by teachers. Based on our observations, we expect that this also have to do with the missing multiplication sign when working with coefficients. When the coefficient is then preceded by a minus sign, some students' schemes might relate to basic calculations working with a subtrahend rather than an isolated or directed number (Gallardo, 2002).

The affordance of the modified balance model becomes the enabling of students to engage with situations that allow for experimentation with the terms as properly directed numbers when solving equations. Experimentations, which allowed the users to see for themselves that a positive term can be seen as pressing down on the balance pan and a negative term, can be seen as pulling up on the balance pan. This directional conception of numbers could be thought of as similar to a force directed by the number's placement on the number line.

Before the teaching experience with Equation Lab, Albert showed in the pretest that he did not have a scheme for solving equations made abstract by negativity. He was seemingly only capable of applying a guess-and-check or counting strategy when solving equations. In the interview 2 weeks after the teaching experience with Equation Lab, Albert demonstrated a strategy that, firstly, helped make inferences of the equations, in terms of relevant concepts and theorems related specifically to the teaching with Equation Lab. The strategy of inverting terms before moving them to the other side, or inverting both sides to make better sense of a coefficient, links directly to the teaching experience with

Equation Lab. In addition, his teacher let us know that he observed Albert apply this strategy of 'inverting' and changing sides in pen-and-paper exercises in the classroom.

Albert was, however, not able to convince us during the interview 2 weeks after the teaching experience with Equation Lab that his equation solving scheme was not adapted to master the situations we presented to him. He did show us how perceptively different the operational minus sign and the minus sign as prefix can be during equation solving. This suggests an emphasis on translating between equations represented on the balance (using prefixes, weights as directional numbers, placing together as addition and balance as equality) and equations represented on the pen-and-paper (using equals signs and mathematical operators) should be the focus of future improvements.

With Albert, we did experience 'detachment of the minus sign' (Herscovics & Linchevski, 1994) in the interview 2 weeks after the teaching experience. When teaching with the modified balance, this phenomenon is definitely something that requires focus. When we worked with Albert during the interview 2 weeks after the teaching experience, he demonstrated to us that the missing operators, when using the modified balance in VR, translate poorly into pen-and-paper. Because the focus of working with Equation Lab was to enable students to work and learn about equations made abstract by negative numbers, the application needs improvements to accommodate this issue. The missing operators on the balance are what were causing this particular issue. Equations such as $-6x = 24$ are where experimentation in the virtual environment shows the most potential. We propose features later in this discussion.

This collectively suggests that teaching with this dynamic modified balance model with its invert mechanic can provide students with a better strategy than guess-and-check and possibly better schemes for solving equations. According to Linsell (2009), having access to more or better strategies for solving linear equations reflects the developmental stage of understanding the concept of equations. The results showed us that it was possible to use the experience with the application to expand the students' sets of equation solving strategies.

We do not claim that a half-hour experience of being taught with Equation Lab makes students' schemes able to be successful in every aspect of linear equation solving. Rather, we note how a student such as Albert can adapt to strategies that likely help him understand more aspects of the concept of linear equations. The strategy of inverting and moving elements (objects, terms) emulates the formal transformations in the case of the balance model to a completely new extent. This is not solely because the modified balance model accepts negative numbers and negativity in this environment, but is also due to the invert mechanic.

To emulate formal transformations fully, as we know them from pen-and-paper exercises, the user would necessarily need the mechanic allowing them to generate new objects. This can be exemplified in the scenario where the task was to solve $7x - 5 = 23$. The user would have been able to spawn two objects, each with a value of 5, to place one on each side, thereby maintaining balance. Afterwards, the student could infer that -5 and 5 would cancel out on the left side. This mechanic, or possibility, of generating elements, is not enabled in the current version. This feature might be a valuable include for the future.

It might be worth discussing how the workflow using the invert mechanic differs from the workflow using a more traditional ‘change side – change sign’ strategy. In the case of an arithmetic equation like $ax = b$, where a is a negative number, let us look at an immediate difference. The invert mechanic allows the user to invert one side of the balance to find that the other side needs to be inverted as well, due to the inverted ‘weight’ of the two terms. The user can infer this by arguing that the invert mechanic changes push to pull, and vice versa, so if one side pushes instead of pulling, the other side needs to do the opposite of what it initially did. Performing this traditional strategy would, in this case, work perfectly fine as well, but the workflow would differ, in the sense that the terms would swap places on the balance or swap sides relative to the equals sign on paper.

This interaction has no immediate meaning attached to it other than fitting the phrase ‘change side – change sign’. In addition, if the users were to apply the push/pull thought process, the added movement to the process could easily confuse them. The invert mechanic might also allow teachers to address or teach additive inverses. One might argue that this discussion does not belong in lower secondary school, but observing how the students used the invert mechanic in Equation Lab has raised our confidence. Research studies such as Vlassis (2002, 2004) emphasise the need to focus more on negative numbers and negativity in general when teaching mathematics. Therefore, our suggestion would be to study additive inverses (Gallardo, 2002), as well as the process of inverting numbers in additional settings not only related to linear equations.

During the interview 2 weeks after the teaching sequence with Equation Lab, Albert demonstrated to us that we need additional focus on the difference between the minus sign as a prefix and as an operator. This means that, even though working with Equation Lab and the mechanics within can benefit the range of strategies available to students, this does not necessarily help them understand how the operational minus sign influences a solution strategy in a pen-and-paper setting. This leads to a discussion of how to improve the environment to assist teachers and, most of all, students, in the quest for understanding the concept of linear equations involving negativity.

When observing the users in Equation Lab, to a great extent we were reminded how much work goes into performing the transformations. Because the user needs to remove a term from the balance, before altering it in the transformation area, finally to collect it from the spawning area (the area opposite the balance where objects appear), we realise that this procedure might be so comprehensive that some experimental value is lost. Because the procedure takes a fair bit of movement and time, the user might lose a bit of focus on what initially was the purpose of altering the term. We imagine a set-up where the alterations of the terms become more flexible. The invert mechanic could work as a transformation tool with the terms still resting on the virtual balance. This would greatly decrease the workload for this specific procedure. This might also be a valuable inclusion for the future. Additionally, the split mechanic and the divide mechanic could perform differently; this would, however, require greater planning, because multiple objects are created from a single one.

To place more focus on how to transfer the use of strategies from Equation Lab to the classroom, we need to focus on the differences and how to overcome them. One obvious difference is representation. The difference in representation is one of the main positives when using the balance model for teaching linear equations. Nevertheless, it is important to teach students about the change in representation and how to change it when moving from Equation Lab or the balance model to pen-and-paper. Instead of leaving this as a task for the teacher to accomplish, we propose the following possible aid to the environment to overcome the issue and to be able to address the 'detachment of the minus sign' within the virtual environment. This could be done by adding a dynamic blackboard to the environment that displays the current state of the balance in mathematical notation (see Fig. 10).

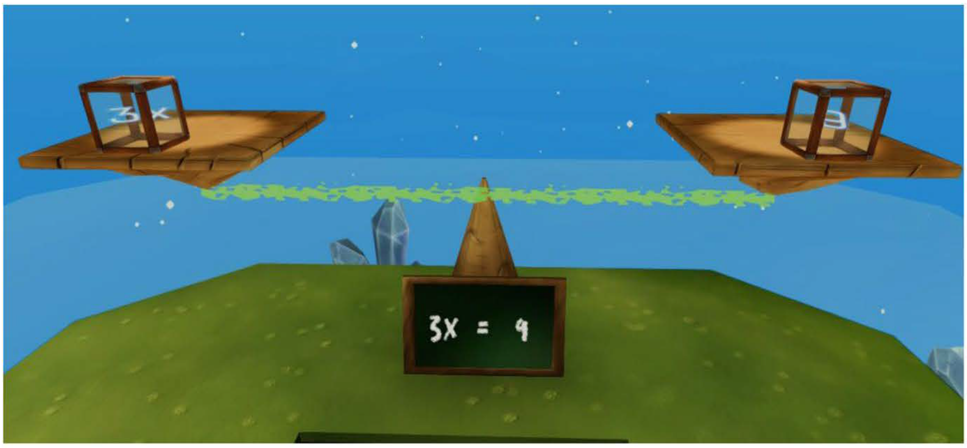


Fig. 10 Dynamic blackboard

The blackboard can display the current state of the balance at any time. When the balance is off or not levelled, the equals sign on the blackboard could change to not equal ($\neq$). An alternative could be that the blackboard displayed 'larger than' or 'smaller than', accordingly, but that could potentially eliminate some exploratory parts of the environment. A further consideration is to have the dynamic blackboard as an optional element in the environment that could be used in teaching situations when needed. Another observation we made was the potential of a possible setting making it possible to choose whether a multiplication sign should be visible or invisible between coefficients and unknowns.

Conclusion

In the introduction, we mentioned a question from Vlassis (2002) about whether we should reject the balance model when teaching equation solving. Based on our design and user study, we believe to have identified positive prospects of teaching equation solving with Equation Lab. With the modification, allowing for the presence of negative numbers, teachers can, in a more complete sense, use the balance model to teach the concept of equations and equation solving strategies. The transformations make sense and feel intuitive to students, since teachers and students can reflect on and discuss the consequences of pressing down on one side and pulling up on the other. Albert highlighted, when solving $-6x = 24$ in Equation Lab, that a dynamic representation of the balance model may help students make inferences that lead to a more familiar situation and possibly to a solution.

Based on students' responses and the case of Albert, we find that learning about negative numbers and how they influence linear equations, utilising the invert mechanic and the modified balance model in this new setting can indeed help to provide situations that potentially strengthen students' schemes related to linear equations. They can get better schemes and conceptual fields regarding equations, adapting to negative numbers as a solution and the goal-oriented transformations of linear equations. However, several design considerations are eligible for further research to understand the effect of representation and interaction on transferability from the new to traditional settings (e.g. a classroom).

Funding This work is possible due to grants from Innovation Fund Denmark, grants 8053-00042A and 7038-00064B.

Data Availability The datasets generated and/or analysed during the current study are not publicly available, due to GDPR.

Declarations

Conflict of Interest The authors declare no competing interests.

References

Ahl, L., & Helenius, O. (2018). Why is calculating the average speed difficult? In E. Bergqvist, M. Österholm, C. Granberg & L. Sumper , *Proceedings of the 42nd Conference of the International Group for the Psychology of Mathematics Education* (2, 3–10). Umeå, Sweden: PME.

Bodin, A. (1993). What does to assess mean? The case of assessing mathematical knowledge. In M. Niss (Ed.), *Investigations into assessment in mathematics education: An ICMI study* (pp. 113–141). Springer.

Falkner, K., Levi, L., & Carpenter, T. (1999). Children's understanding of equality: A foundation for algebra. *Teaching Children Mathematics, 6*(4), 232–236.

Filloy, E., & Rojano, T. (1989). Solving equations: The transition from arithmetic to algebra. *For the Learning of Mathematics, 9*(2), 19–25.

Gallardo, A. (2002). The extension of the natural-number domain to the integers in the transition from arithmetic to algebra. *Educational Studies in Mathematics, 49*(2), 171–192.

Herscovics, N., & Linchevski, L. (1994). A cognitive gap between arithmetic and algebra. *Educational Studies in Mathematics, 27*(1), 59–78.

Jankvist, U., & Niss, M. (2017). The notion and role of "detection tests" in the Danish upper secondary "maths counsellor" programme. In T. Dooley & G. Gueudet (Eds.), *Prooceedings of the Tenth Congress of the European Society for Research in Mathematics Education* (pp. 3825–3832). Institute of Education, Dublin City University and ERME.

Kieran, C. (1981). Concepts associated with the equality symbol. *Educational Studies in Mathematics, 12*(3), 317–326.

Kieran, C. (1992). The learning and teaching of school algebra. In D. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 390–419). Macmillan Publishing Company.

Kieran, C. (2007). Learning and teaching algebra at the middle school through college levels: Building meaning for symbols and their manipulation. In F. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 707–762). Information Age Publishing.

Linsell, C. (2009). A hierarchy of strategies for solving linear equations. In R. Hunter, B. Bicknell & T. Burgess , *Crossing divides: Proceedings of the 32nd annual conference of the Mathematics Education Research Group of Australasia* (1, 331–338). Palmerston North, NZ: MERGA.

Matthews, P., Rittle-Johnson, B., McEldoon, K., & Taylor, R. (2012). Measure for measure: What combining diverse measures reveals about children's understanding of the equal sign as an indicator of mathematical equality. *Journal for Research in Mathematics Education, 43*(3), 316–350.

Moyer-Packenham, P., & Bolyard, J. (2016). Revisiting the definition of a virtual manipulative. In P. Moyer-Packenham (Ed.), *International perspectives on teaching and learning mathematics with virtual manipulatives* (pp. 3–23). Springer.

Otten, M., van den Heuvel-Panhuizen, M., & Veldhuis, M. (2019). The balance model for teaching linear equations: A systematic literature review. *International Journal of STEM Education, 6*(1), 30–30.

Pirie, S., & Martin, L. (1997). The equation, the whole equation and nothing but the equation! One approach to the teaching of linear equations. *Educational Studies in Mathematics, 34*(2), 159–181.

Rhine, S., Harrington, R., & Starr, C. (2018). *How students think when doing algebra*. Information Age Publishing.

Suh, J., & Moyer-Packenham, P. (2007). Developing students' representational fluency using virtual and physical algebra balances. *Journal of Computers in Mathematics and Science Teaching, 26*(2), 155–173.

Tall, D., & Vinner, S. (1981). Concept image and concept definition in mathematics with particular reference to limits and continuity. *Educational Studies in Mathematics, 12*(2), 151–169.

Vergnaud, G. (1988). A comprehensive theory of representation for mathematics education. *The Journal of Mathematical Behavior, 17*(2), 167–181.

Vergnaud, G. (1997). The nature of mathematical concepts. In T. Nunes & P. Bryant (Eds.), *Learning and teaching mathematics: An international perspective* (pp. 5–27). Psychology Press.

Vergnaud, G. (2009). The theory of conceptual fields. *Human Development, 52*(2), 83–94.

Vlassis, J. (2002). The balance model: Hindrance or support for the solving of linear equations with one unknown. *Educational Studies in Mathematics, 49*(3), 341–359.

Vlassis, J. (2004). Making sense of the minus sign or becoming flexible in 'negativity.' *Learning and Instruction, 14*(5), 469–484.

Appendix 1: Supplementary – Figures and Tables from MIRM(100)

Corresponding results considering only students who answered at least 100 questions. This reduced data set contains only 2169 students.

Most common answer	Most common student answer	Most common question answer	Most probable MIRM answer
81.89 % accuracy	81.89 % accuracy	82.13 % accuracy	82.19 % accuracy

Table 1s: Answer prediction results leaving 1% of answers as missing during inferencing predicted by i) overall most common answer type, ii) most common student answer, iii) most common question answer and iv) most probable answer according to the MIRM. The McNemar test identified no significant differences between the approaches.

Appendix 2: Questions from the survey on teachers' use of statistical tools in MatematikFessor

The survey consisted of the following questions:

1) How often do you use MatematikFessor in your teaching?
    a. Often
    b. Sometimes
    c. Never or almost never
2) Do you use the tool 'Live statistics'?
    a. Yes
    b. No
    c. I did not know such a tool existed.
3) Do you use the tool 'Student/task statistics'?
    a. Yes
    b. No
    c. I did not know such a tool existed.
4) Do you use the tool 'Topic statistics'?
    a. Yes
    b. No
    c. I did not know such a tool existed.
5) Do you use the tool 'Complete/school to home statistics'?
    a. Yes
    b. No
    c. I did not know such a tool existed.

Appendix 3: Co-clustering as a statistical tool

In this section, I provide a short theoretical overview of the machine learning techniques utilised in this project. The main goal of machine learning algorithms is to learn from past experiences (observations) to generalise future situations of the same type with as little human intervention as possible. In general, there exist two areas of machine learning: supervised learning and unsupervised learning. Supervised learning is a method for establishing a model based on already known input and output pairs. Such a model can be used to solve classification or regression problems. Unsupervised learning is used to establish a categorisation (clustering) of the input data. Unsupervised learning is used to establish a model that can cluster data based on similarities, for instance, creating a model that groups pictures based on what animal is in the picture. A primitive model might be able to distinguish large animals from small animals, and a better model might be able to distinguish animals based on where they are found in the wild or those able to fly from those that cannot (Herlau et al., 2022).

Co-clustering is when the machine attempts to learn about groups along several axes. In the case of this project, I worked with students to solve tasks in an online learning environment. As I explained in further detail later, I sought to generate knowledge about the students as well as the tasks they engaged with in the online learning environment.