Hypothesis generation: How to put students into motion?
HYPOTHESIS GENERATION:
HOW TO PUT STUDENTS INTO MOTION?

Xiulin Kuang
况秀林
HYPOTHESIS GENERATION:
HOW TO PUT STUDENTS INTO MOTION?

DISSERTATION

to obtain
the degree of doctor at the University of Twente,
on the authority of the rector magnificus,
prof. dr. ir. A. Veldkamp,
on account of the decision of the Doctorate Board,
to be publicly defended
on Friday the 7th of December 2022 at 12:45 pm

by

Xiulin Kuang
况秀林

born on the 5th of February 1990
in Wuhan, the People’s Republic of China
Graduation committee

Chair: Prof. dr. T. Bondarouk – University of Twente
Supervisor: Prof. dr. A.J.M. de Jong - University of Twente
Co-supervisor: Dr. T.H.S. Eysink - University of Twente

Committee members: Prof. dr. S.E. McKenney - University of Twente
Prof. dr. J.H. Walma van der Molen - University of Twente
Prof. dr. Z. Zacharia - University of Cyprus
Prof. dr. C.Y. Chang - National Taiwan Normal University
Dr. P. Papadopoulos - University of Twente
The research reported in this thesis was funded by the China Scholarship Council (grant number: 201707720027).
Acknowledgements

“Who would you like to thank during your PhD journey?” is probably the easiest PhD-related question. If this was one question of my upcoming thesis defense, I am more than confident that I can pass. Thinking of this question allows me to replay the memories of my PhD life, and many people who directly or indirectly helped me come to my mind. This thesis and all the work done towards it would not be possible without the contribution and support of many people. Here, I would like to mention some individuals in particular.

I would first give my greatest thanks to my supervisors Ton de Jong and Tessa Eysink. Ton, many thanks for always being reachable and responsive to my questions throughout both my master and PhD years, for gently pushing me to make decisions and to grow as an independent researcher, for providing knowledgeable feedback and constructive criticism on my work, for magically simplifying the big issues that I struggled with, for encouraging me to ask for help when I needed, and for congratulating on all the mini- or big milestones throughout my project. I would also like to express my sincere gratitude to Tessa. Thank you for meeting me regularly to provide continued support, for providing all the detailed and insightful suggestions on my work, for being a good listener and problem solver to all my questions and doubts, for letting me know that you are always there to help either in work or in life, and for sending me all the sweet cards and wishes during the Christmas time.

I would also like to thank the China Scholarship Council for funding my PhD project, which gave me the chance to be on board. Thanks for choosing me and believing in me, which support me both financially and mentally.

I have enjoyed doing my PhD at the University of Twente, and this is probably mainly because of the positive and supportive working environment in our IST
department. Big thanks to all my colleagues of the IST. Thank you all for sharing inspiring ideas and insights in colloquiums and research meetings, for always willing to help both academically and personally, for introducing me to Dutch food during lunch time, for translating Dutch game rules and menus for me during ‘IST uitjes’, and for sharing traditional or self-made snacks. Sandra, thank you for helping me with all the administrative issues, for making the office more fun with festival decorations, for surprising me with birthday decoration in my office, and especially for kindly helping out during the final phase of my PhD. Jakob, thank you for all your support in the tool design and troubleshooting for my studies, for all your invitations for a lunch walk, and for all the unique new year cards. Henny and Ellen, thanks for your supervision and guidance on my Master thesis, which built a good starting point for my PhD journey. Hagop, thank you for your technical support on realizing a new tool for my study.

Thank you also goes to my fellow PhD students, who have made this journey more inspiring and exciting. I really enjoyed the time that we spent together, in the office, in the ProIST, in the Theater Café, in the Sinterklaas activity. I would especially thank my roommate Natasha, the care and support you have shown mean so much to me. You are always the one that I can go to when everything seems unclear. It feels really good to have a friend in the office. Thanks for all the laughs, advice, guidance, encouragements, gifts, snacks, and wishes all along the journey. Elise, I always remember how I was touched when you invited me to go to the ICO course together and showed me the way and location of the lecture room in Utrecht for the first time. Thank you for your kindness and willingness for help throughout time. Johannes, thank you for all the clear explanations for my statistical questions. Irena, thank you for sharing the same mindset as me about how to manage the housing things alone. And Karel, Luiza, Sharanya, Adelson, Quirine, thank you for being around to help with language translation, to offer input on my draft versions of work, and to listen to the challenges that I have in my studies.

I would also like to express my sincere appreciation to the teams who designed, developed and supported Graasp platform and Go-lab ecosystem, with which all learning environments of my studies were designed. Thank you for offering the amazing user-friendly platform, for providing a large collection of open educational resources like simulation labs, scaffolding apps, and authorizing and sharing tools. Denis, André, and Juan Carlos, thank you for your timely assistance in the troubleshooting. I really enjoyed designing and sharing my ILSs on Graasp.
I have also relied on the help and support of many people outside our department. Emily, thanks for your thorough edits and critical comments, which helped to make the papers more understandable and clear. Effie, Matthias, Pamela, thank you for your warm welcome when I came to Britain for my experiment, and thank you for your kind assistance during my data collection. I am also thankful to all the physics teachers from different countries that were involved in my studies. Daniela, Lorena, Adrian, Marisa, Patrick, David, Lidia, Simona, Jan, Yara, thank you for offering me the opportunities and support to carry out my experiments with your students. Thank you for translating the ILSs into local languages, for discussing and checking the knowledge tests, for arranging the computer rooms and facilities for my experiments, and for being supportive and cooperative in helping me to complete my experiments.

To my parents and brother, thank you for your support and love. Sometimes, it was not easy for me to talk about what was going on when we were far away from each other, but thank you for always being there at the other end of the line. Knowing you are always by my side was a huge source of comfort and confidence during the journey.

To my parents-in-law, thank you for always being kind and supportive throughout the years, especially this year. Thank you for sharing the load of motherhood, for taking good care of Kexiao and saving me several hours a day to work on this thesis. Thank you for all the sacrifices you have made in your personal life to support me.

To my lovely son, Kexiao, thank you for pushing me to grow up as an ‘early bird’, for lighting up my tough days with your smile, for keeping me feel fulfilled, and for sharing your toys with me for the inspiration for the cover page.

Finally, to my husband, Ye, thanks for being the one that I can always talk to and rely on. Thank you for sharing all my ups and downs along the journey, for always listening to me and putting up with me, for helping me to navigate through my struggles, for cooking some Chinese food or bringing back some flowers to encourage me, for pushing me to go out for a walk or cycling to cheer me up, for celebrating each milestone with a simple text or a big dinner, for taking charge of housework when deadlines approached, for all the little and big things you did to make this PhD journey easier and happier.

Xiulin
Table of contents

Chapter 1: General introduction .......................................................... 1
  Introduction ................................................................................. 2
  Inquiry learning ....................................................................... 2
  Hypothesis generation ............................................................ 3
  Learning environment and domain ............................................. 4
  Problem statement and dissertation outline .............................. 6
  References .................................................................................. 8

Chapter 2: Providing partial hypotheses as support .......................... 13
  Abstract ..................................................................................... 13
  Introduction .............................................................................. 14
  Method ...................................................................................... 20
  Results ....................................................................................... 33
  Discussion ................................................................................ 39
  References ............................................................................... 42

Chapter 3: Providing domain information and/or self-exploration as support ................................................................. 49
  Abstract ..................................................................................... 49
  Introduction .............................................................................. 50
  Method ...................................................................................... 56
  Results ....................................................................................... 68
  Discussion ................................................................................ 73
  References ............................................................................... 77
## Chapter 4: Providing adaptive domain information as support

- Abstract .................................................. 85
- Introduction .............................................. 86
- Method ...................................................... 91
- Results .................................................... 101
- Discussion ................................................. 104
- References ................................................ 107

## Chapter 5: General discussion

- Introduction ............................................. 114
- The design of the studies and main findings ........ 115
- Discussion ............................................... 117
- General conclusion .................................... 125
- References ............................................... 125

**English summary** ........................................ 129

**Nederlandse samenvatting** ........................... 135

**Appendices** ............................................. 141

- Appendix A .............................................. 142
- Appendix B ............................................... 144
- Appendix C .............................................. 146
Chapter 1

General introduction
Introduction

Imagine a student thinking about why a large wooden stick can float on the water while a small stone will sink into the water. The answer cannot be found by randomly throwing sticks, and stones in the water. Instead, the student needs to organize initial ideas about possible explanations and then test each of them by deliberately choosing sticks and stones to throw in the water and observe. The more prior domain knowledge a student has about the relevant characteristics of sticks, stones, and water (like volume, mass, temperature, and density), the more complete the answer that the student is able to end up with. This scenario reflects the main idea of hypothesis generation, which is a process to organize what you know about a domain to make sense of a problem and to organize tentative answers as testable propositions to guide further investigation (de Jong & van Joolingen, 1998; Kim & Pedersen, 2011). Hypothesis generation is a crucial but challenging process for students (Chang et al., 2008; de Jong & van Joolingen, 1998; Kim & Pedersen, 2011; Njoo & de Jong, 1993). This dissertation explores ways to support secondary school students’ hypothesis generation performance.

Inquiry learning

Hypothesis generation is one of the processes of inquiry learning. The main idea of inquiry learning is grounded on the work of John Dewey (1859-1952), who in the early 20th century advocated that science education should be more about learning through experiences rather than memorizing scientific facts. Inspired by his work, there has been a growing recognition of inquiry learning, and inquiry learning theory has been modified and enriched over the years. There are many variations in the definition of inquiry learning, but in essence, inquiry learning emphasizes the involvement of students in the learning process and the active construction of domain knowledge and acquisition of scientific skills as the learning goals (de Jong, 2019; Edelson et al., 1999; Lazonder & Harmsen, 2016). In this dissertation, inquiry learning is defined as a student-centered learning activity to provide personal experiences for students to understand science content and also scientific practices. An inquiry activity often starts with a question or a problem, which then is followed by, but not limited to, generating hypotheses, designing experiments, collecting data, and drawing conclusions to find out answers or solutions (de Jong, 2006; Pedaste et al., 2015).
However, participation in such a student-centered learning activity does not mean leaving students all on themselves to complete the inquiry processes. Taking initiatives in inquiry learning requires much cognitive ability and skills, which many students may lack. Students may come across various difficulties during inquiry activities (de Jong, 2006; Ferrés Gurt & Marbà Tallada, 2018) and, therefore, support is needed for students to benefit from inquiry learning (Alfieri et al., 2011; Belland et al., 2017; D’Angelo et al., 2014; Lazonder & Harmsen, 2016). Indeed, many studies have shown the necessity of providing support for students in inquiry learning. A meta-analysis from Alfieri et al. (2011) found superior outcomes of supported inquiry learning, but inferior outcomes of unsupported inquiry learning compared with other forms of instruction. Lazonder and Harmsen (2016) confirmed that providing support has a significant positive effect on inquiry learning activities, performance success, and learning outcomes.

Different types of supports have been promoted and examined to facilitate students’ performance in inquiry learning. These are face-to-face supports provided by human tutors, such as presentation of background domain knowledge before inquiry learning and subsequent presentation of theoretical ideas at the end of inquiry learning through teacher-led introduction and discussion (Wecker et al., 2013). Or supports by fellow students in collaborative inquiry learning such as peer feedback (Dmoshinskaia et al., 2021). Or digital supports provided by online learning environments such as performance dashboard (Hagemans et al., 2013), prompts (Kim & Pedersen, 2011), automatic feedback (Kroeze et al., 2019), and a scaffolding tool to design experiments (van Riesen et al., 2018). This dissertation focuses on digital support within the learning environment. This is because preserving students’ autonomy in inquiry learning ideally requires teachers to cater students’ individual needs in time, which is hardly possible in classroom settings. Digital support, if designed appropriately, may help to satisfy individual students’ active inquiry experiences.

Hypothesis generation

Hypothesis generation is regarded as an important process of inquiry learning (Gyllenpalm & Wickman, 2011; Kim et al., 2009; van Joolingen & de Jong, 1991). Inquiry learning emphasizes the active engagement of students in constructing a conceptual model or rule about the essential features of a given domain (van Joolingen & de Jong, 1993). It can even be regarded as a rule discovery task. But
when leaving students to discover the rule all on themselves, they often do so with mindless searching or exploring (Klahr & Dunbar, 1988; Mayer, 2004). One way to trigger a general awareness of the to-be-solved problem and prepare students with to-be-tested rules is to ask them to write hypotheses. Hypothesis generation is a thought-provoking and thought-organizing process. The generation of hypotheses requires students to relate the to-be-solved problem to relevant knowledge, and organize the outcome as hypotheses, or relations between variables. The relations between variables then act as references on what to consider in experiment design, what to observe and record in data recording, and what to conclude in drawing conclusions. The testing outcome of each generated hypothesis would then help to form students’ understanding of the domain.

Despite its importance, hypothesis generation is a challenging process for students (Chang et al., 2008; de Jong & van Joolingen, 1998; Kim et al., 2009; Njoo & de Jong, 1993). Students may not know the main elements of a hypothesis or what a hypothesis looks like (de Jong & van Joolingen, 1998), or they find it hard to decide which items to focus on from a scientific point of view (Bell et al., 2010). Another problem is that they have difficulties in generating informative and testable hypotheses (de Jong & van Joolingen, 1998; Gijlers & de Jong, 2009). Some students, especially students with low level of prior knowledge, have problems in identifying variables and indicate potential relations between variables (de Jong & van Joolingen, 1998), and it is hard for them to generate alternative hypotheses (Kim & Pedersen, 2011). Hence, it is of significance to support students’ hypothesis generation.

**Learning environment and domain**

All studies in this dissertation were conducted with inquiry learning activities in regular classroom settings with technology-based learning environments. The first reason to focus on technology-based learning environment is that compared with traditional lectures, computer technologies help to create more manageable learning environments for inquiry learning in classroom settings (van Joolingen et al., 2007). Inquiry learning requires a large amount of preparation work either by students or tutors, such as documents on relevant domain information to consult, experimental apparatus and laboratory supplies to carry out experiments, and notebooks to record and analyze data. Computer technologies can simulate and integrate all this work in one learning environment, which helps to reduce some...
workload for students in participating in inquiry learning and also for teachers in supporting their students in class. The second reason is that computer technology can provide affordances that physical laboratories cannot offer (Olympiou et al., 2013; Pedaste et al., 2015). For instance, technology-based learning environments can simulate some phenomena or variables that are impossible in regular classrooms, such as friction-free ground and the motion of microscopic particles. Besides, technology-based learning environments can offer cognitive tools to support students’ inquiry learning instantly based on their needs.

The online inquiry learning environments used in the three studies of this dissertation were all designed based on the Go-Lab ecosystem, which is an open online design and sharing platform for science teaching (de Jong et al., 2021). It offers inquiry learning templates, simulation labs, cognitive supporting tools for teachers to choose, customize, and assemble to create an inquiry learning environment.

More specific for the studies in this dissertation, the hypothesis scratchpad, aiming to facilitate generation of hypothesis, was the major tool students used to write their hypotheses. The first version of the hypothesis scratchpad was designed by van Joolingen and de Jong (1991). Considering that generating a hypothesis mainly includes three sub-processes (identifying variables, selecting variables, and defining relations between the selected variables) they provided three tables to support these sub-processes. These three tables respectively contained a list of variables and information on the variables, a list of conditions, and a list of relations for students to choose from to complete their hypotheses. They compared the effectiveness of providing all the tables or some of them in facilitating hypothesis generation and found out that the fully structured hypothesis scratchpad with three tables was effective in supporting the generation of well-formed hypotheses. The hypothesis scratchpad used in the present dissertation was a more concise and flexible version, which has two major parts: a term section shown with terms on the three main elements of a hypothesis (variables, conditions, and relations) on a learning domain and an extra editable and reusable term block, and a hypothesis writing section to drag and drop the given terms to build up hypotheses (see Figure 1.1). In the studies of this dissertation, we configured the hypothesis scratchpad according to the research goal of each included study. The variable terms covered all the variables in the simulation lab, which was to be used for experimentation to test hypotheses in a target domain. The terms on relations and conditions were also deliberately chosen to cover most potential situations. Students were also allowed to use their own terms.
Students in the studies of this dissertation worked on a physics domain – force and motion – with the online inquiry learning environments. To be more specific, students were supposed to find out the relationship between force and motion. This topic was chosen for two reasons. First, students are prone to have misunderstandings on the relationship between force and motion. Based on daily experience, it is easy to imagine that forces can change motion. But quite a few students know motion can be there without force, which makes this topic interesting to explore. Second, this topic fitted in the curriculum of the participating classes as this was arranged with their teachers. To make the inquiry learning method more feasible in classroom settings, the learning topic was chosen together with the physics teachers of the target schools to fit in the normal teaching plan.

Problem statement and dissertation outline

As mentioned earlier, hypothesis generation is an important but difficult learning process which needs to be supported. The major goal of the studies in this dissertation was to explore ways to support students’ hypothesis generation in online inquiry learning. Three kinds of support were developed and examined in the three studies of this dissertation. A defining difference between inquiry learning and traditional learning is the freedom that is given to students to explore and investigate a learning domain by themselves. One of the critical tasks for researchers in the education field is to balance between the amount of support and
the freedom allowed for students, and to provide just enough support (Belland et al., 2017; de Jong, 2019; Lazonder et al., 2009). The studies of present dissertation were explorations to find this balance step by step.

**Study 1**

This study examined the effect of partial directive support by providing partial hypotheses. Directive support facilitates students by steering them in a certain direction, which limits their freedom (de Jong & Lazonder, 2014; Njoo & de Jong, 1993). The idea to provide partial hypotheses as support was inspired by the completion strategy, which provides partial solutions to direct students’ attention to more productive parts of a learning task (van Merriënboer, 1990). The support is called partial directive support since students were allowed to think of and complete the second half of the hypotheses. The effect of the partial hypotheses was investigated by comparing two conditions. Students were provided with either a set of terms on variables, relations and conditions, or the same set of terms and additionally a partial hypothesis to write their hypotheses with.

**Study 2**

Considering that a given partial hypothesis still limits students’ freedom to express their own thoughts in hypotheses, in Study 2, more non-directive support was provided for students. In contrast with directive support, non-directive support aims to help students in fulfilling the ideas that they have in mind (de Jong & Lazonder, 2014; Njoo & de Jong, 1993). In this study, we focused more on finding out the right way to prepare students with domain knowledge, so that they could be able to generate their own hypotheses. The two knowledge sources for hypothesis generation suggested by the SDDS (scientific discovery as dual search) model of Klahr and Dunbar (1988) guided the design of the support in this study. According to this model, students generate their hypotheses either based on their prior knowledge or based on their experiences from some prior exploratory experiments. Hence, we investigated whether providing domain knowledge together with exploratory practice or providing either of them separately can facilitate students’ performance on hypothesis generation.
Chapter 1

Study 3

The ultimate goal of providing support is to give students what they need. A limitation of the first two studies was that the given support, if found effective, would be assumed to benefit all students. To provide better support for students, and to provide better choices for teachers, in Study 3, individual differences in students’ prior knowledge was taken into consideration to provide more individually adaptive domain information as the support.

The effects of all the above-mentioned types of support were examined on students’ knowledge acquisition and performance on hypothesis generation. Since the generated hypotheses were an important basis for subsequent inquiry learning processes, students’ performance on processes further on in the inquiry cycle, such as data recording and drawing conclusions was also included in the data analysis. The three studies are presented in Chapters 2 to 4 respectively. General conclusions and a discussion of the findings and methodologies of all studies is presented in Chapter 5.

References


Chapter 1
Chapter 2
Providing partial hypotheses as support

Abstract

Hypothesis generation is an important but difficult process for students. This study investigated the effects of providing students with support for hypothesis generation, with regard to the testability and complexity of the generated hypotheses, the quality of the subsequent inquiry learning processes, and knowledge acquisition. Fifty-two secondary school students completed three prior knowledge tests and worked on an inquiry task in the domain of force and motion, concerning the topic of Newton’s first law of motion. They received either a set of terms (variables, conditions, and relations) to help them generate hypotheses (T condition, n = 23) or the same set of terms plus a partial hypothesis to start from (T+Phy condition, n = 29). Results showed that students in the T+Phy condition generated more complex hypotheses, performed better at data collection, and acquired more domain knowledge than students in the T condition. No effects of prior knowledge were found.

* This chapter is based on:
Introduction

Inquiry learning is an instructional method for science education that emphasizes the active involvement of students. In an inquiry context, students are expected to actively explore problems or phenomena in a way that resembles what scientists do — asking questions, generating hypotheses, designing experiments, and drawing conclusions. Computer simulations, or online laboratories, are receiving increasing attention from the educational field because of their potential to provide a feasible learning environment for inquiry learning (Blake & Scanlon, 2007; Lai, Hwang, & Tu, 2018; Rieber, Tzeng, & Tribble, 2004; van Joolingen & de Jong, 1991). Compared with hands-on inquiry learning, simulation-based inquiry learning has the advantages that it can provide a relatively harmless inquiry environment for students (de Jong, 2006), can make normally invisible phenomena visible (Olympiou, Zacharias, & de Jong, 2013; Windschitl, 2000), and can simplify or emphasize certain aspects of the domain to facilitate students’ understanding of the phenomenon (van Joolingen, de Jong, & Dimitrakopoulou, 2007).

Despite the widespread belief that simulation-based inquiry learning is a promising learning method (de Jong, 2006; Rutten, van Joolingen, & van der Veen, 2012; Woolf et al., 2002), use of simulation-based inquiry does not by itself guarantee effectiveness in facilitating learning (de Jong & van Joolingen, 1998; Eslinger, White, Frederiksen, & Brobst, 2008; Keselman, 2003). Empirical studies have found that students often lack the inquiry skills needed to complete systematic scientific investigations (Arnold, Kremer, & Mayer, 2014; Edelson, Gordin, & Pea, 1999; Krajcik et al., 1998). It is, therefore, widely agreed that inquiry with minimal guidance or even no guidance is less effective and less efficient than guided inquiry learning (Hmelo-Silver, Duncan, & Chinn, 2007; Kirschner, Sweller, & Clark, 2006; Mayer, 2004), which has also been confirmed in a number of overview studies (Lazonder & Harmsen, 2016; Rutten et al., 2012; Smetana & Bell, 2012).

Importance and difficulty of hypothesis generation

Students can encounter problems during different inquiry learning processes, and dedicated supports are needed for these processes. The current study focused on support for the hypothesis generation process in a simulation-based inquiry learning environment. The first reason for this focus is that this process and its product are crucial for the entire inquiry learning process. The hypothesis
generation process activates students’ prior knowledge and forces students to mindfully consider variable selection as well as the relationships between the variables involved in the domain (Windschitl & Andre, 1998). The tentative relations between variables stated in the hypotheses that are generated then serve to direct and guide the process of experiment design (de Jong, 2006; Swatton, 1992). During experimentation, hypotheses can direct the data collection (de Jong, 2006; Klahr & Simon, 1999; Swatton, 1992). After completion of the data collection, hypotheses can guide interpretation of data and drawing of conclusions (Piekný & Maehler, 2013).

The second reason is that generating a testable hypothesis is commonly regarded as a difficult process for students (Chang, Chen, Lin, & Sung, 2008; Gijlers & de Jong, 2009; Keselman, 2003; Njoo & de Jong, 1993; Swatton, 1992). A testable hypothesis is a statement that indicates a relation between two or more relevant variables and that is falsifiable. The three main elements of a testable hypothesis are variables, conditions of the variables and relations between the variables (van Joolingen & de Jong, 1991). One often-mentioned problem in hypothesis generation is that students do not know what a testable hypothesis should look like (de Jong & van Joolingen, 1998; Swatton, 1992). Besides their difficulties with testability, students find it hard to identify relevant variables that require investigation (de Jong, 1991), they have problems in differentiating between dependent, independent, and controlled variables, and they frequently fail to identify potential relations between variables (Njoo & de Jong, 1993).

**Support for hypothesis generation**

In spite of the importance and difficulty of hypothesis generation, a limited body of studies has focused on supporting this process. Kim and Pedersen (2011) investigated the effectiveness of metacognitive supports for strengthening hypothesis generation. In their study, students in the treatment condition had access to three types of metacognitive supports during the hypothesis generation process, including reflective prompts, self-questioning, and a self-checklist. All of these supports were aimed at reminding students to reflect on their hypothesis generation process. The results revealed that the presence of the three metacognitive supports together promoted students’ hypothesis development performance and domain knowledge acquisition.
Interpretative support was investigated by Reid, Zhang, and Chen (2003) as another way to help students with their hypothesis generation. In their study, multiple-choice questions and concept descriptions provided prior to the inquiry process were the main elements of the support offered, with the aim of activating students’ relevant knowledge, prompting students to make a general analysis of the problem, and providing a knowledge base for students. The effectiveness of this interpretative support was confirmed in the sense that students who received this support had better conceptual understanding of the domain, performed better on transferring the knowledge to new situations, and associated their prior knowledge better with the rules they discovered (Reid et al., 2003; Zhang, Chen, Sun, & Reid, 2004).

The above two types of supports are less directive forms of support that aim at prompting or preparing the students to state their own hypotheses; there is another, more directive type of support that focuses more on hypothesis generation itself. Based on the consideration that students often do not know what a testable hypothesis looks like, this type of directive support takes the form of a hypothesis scratchpad, initially designed by van Joolingen and de Jong (1991), which aims to facilitate students’ hypothesis generation by providing the hypothesis elements or structure for students to use in completing their hypotheses. This type of support has been further developed into two supporting measures with different levels of directiveness. One of these (first version of the hypothesis scratchpad) provides students with a template giving the three main elements (variables, conditions, and relations) of a hypothesis (van Joolingen & de Jong, 1991). Students can write their own hypotheses by filling in and combining these three elements. The results of the empirical study using this version showed that with the support of the given template, students used a larger number of different variables in their hypotheses (van Joolingen & de Jong, 1991). Another more directive form of support is to offer students a specific number of ready-made hypotheses to use in their inquiry process (Njoo & de Jong, 1993). The results here revealed that students who were provided with pre-defined hypotheses got higher scores for valid learning processes and completed more of the given tasks (Njoo & de Jong, 1993).

Despite their identified advantages, these directive supports were also shown in the studies cited above to have limitations. One is that neither form of support resulted in better performance on domain knowledge acquisition. Another limitation is that the completely pre-defined hypotheses limited students’ freedom
to express their own ideas and forced students to inquire in certain directions. A question therefore arises: how can we optimize the hypothesis scratchpad to take into account both the effectiveness of the support for facilitating students’ knowledge acquisition and students’ amount of freedom? What makes inquiry learning differ from traditional instruction is that students are supposed to be actively engaged in the learning process. Students need enough freedom to become cognitively active in the process of sense making (Mayer, 2004). Although not specifically about inquiry learning, an empirical study on encouraging online active learning found that offering students with instructor-assisted freedom of choice motivated students to actively engage and achieve more in their online studies (Radenski, 2009). There is strong agreement that when designing a support for inquiry learning, it is important to balance between the amount of guidance provided and the amount of freedom allowed for (Bell, Urhahne, Schanze, & Ploetzner, 2010; Njoo & de Jong, 1993). Considering all these, a promising direction for optimizing the effectiveness of the hypothesis scratchpad could be to provide students with a partial sentence that begins the statement of a testable hypothesis, leaving students to complete the hypothesis.

Providing partial solutions as supports

Offering partial hypotheses is related to what van Merriëenboer (1990) called the completion strategy. A completion strategy uses a special form of a worked-out example, in which a partial solution is provided that must be completed by the students. The partial solution limits the task the students need to do, potentially directing their attention to more productive parts of the task (Paas, 1992). In this way, the partial solution that is offered reduces the students’ extraneous cognitive load, which is not essential for attaining the learning goals. The partial solution given also contains important information that can enable students to proceed with the learning process, especially those who lack the necessary schemas or skills (Frerejean, van Strien, Kirschner, & Brand-Gruwel, 2016). With regard to hypothesis generation, the partial hypotheses given to students in this study include a stated condition of an independent variable, which may direct students to figure out a related dependent variable and a relationship between them. It was assumed that the given half-hypothesis can provide a referent example for students to start their hypothesis generation. For instance, when asking about the factors that can affect friction, the given half-hypothesis could be: If the mass of an object increases, then__. With such a support, there is a good likelihood that a student can follow the logic and structure to complete a testable and informative hypothesis. Moreover,
by allowing students the freedom to express their own ideas when completing the incomplete parts, it does not constrain students too much.

The effectiveness of the completion strategy has been examined in a number of empirical studies. Van Merriënboer (1990) used the completion strategy as an intervention for an introductory computer programming course for high school students. Compared with designing and coding a complete program from scratch, the completion strategy involved providing the students with completion assignments that consisted of a problem specification and a partial program to be modified or completed. The results showed that use of the completion strategy effectively facilitated students' use of the programming templates as well as improved their learning and transfer performance. These findings were replicated in a study by Paas (1992), who found that compared with traditional instruction of basic statistics for secondary school students, both partly and completely worked-out examples were less effort-demanding and led to better transfer performance on solving additional problems that differed from the problems presented during instruction.

The completion strategy has also been used in studies in which students had to create a concept map or a model (Chang, Sung, & Chen, 2001). In that case, an expert concept map with some nodes and links reserved as blanks was offered to the students, leaving them to complete the framework by filling in the blanks. This partial support was assumed to reduce mental load and to provide a referent knowledge structure for novice students. Results indicated that this kind of partially completed concept map led students to produce more accurate concept maps, and had a more positive effect on students’ learning than generating a concept map from scratch. More recently, Mulder, Bollen, de Jong, and Lazonder (2016) investigated the effects of partially worked-out models. They compared the effectiveness of two kinds of partial supports versus building models completely from scratch. One was providing students with the overall structure of the model, the other was providing both the structure of the model and a list of variables that should be included in the model. Results demonstrated that students in both partial model conditions built better models, performed better model testing activities, and gained more knowledge than those in the no support condition. In addition, the more extensive support (partial model + variable list) improved students’ knowledge acquisition, model quality, and model testing activities more as compared to the support in which only the partial model was given.
Impact of prior knowledge

Prior knowledge is a factor that has been shown to have an effect on hypothesis generation. A hypothesis holds a student’s tentative idea of the relations between the variables in a domain. Hypothesis generation relies on activating students’ prior knowledge and mapping this onto the problem or question to be addressed (Reid et al., 2003). Students’ prior knowledge of a domain can prepare them with a knowledge base from which to select relevant variables, and familiarity with variables may invoke probable inferences about the relationship between variables. Lavoie and Good (1988) found that learners’ prior knowledge of a domain can help them to generate more accurate hypotheses concerning the relationship between the independent and dependent variables in the domain. Having prior knowledge of the domain could also benefit students as far as generating better hypotheses. Lazonder, Wilhelm, and Hagemans (2008) compared students’ hypothesis generation performance when completing a concrete task and an abstract task. The results demonstrated that learners knew more about the relations between the variables in the concrete task than the abstract task, and their initial hypotheses for the concrete task were more specific than those for the abstract task.

Although not specifically focused on the hypothesis generation process, many studies have found that prior knowledge affects the effectiveness of inquiry learning, and most of these studies have focused on its impact on domain-specific knowledge (Kalyuga, 2008; van Riesen, Gijlers, Anjewierden, & de Jong, 2018; Wang, Wang, Tai, & Chen, 2010). Domain-specific knowledge, conceptual knowledge about a specific topic within the larger domain, is the target type of knowledge to be increased through the inquiry learning process as a whole. But it is not the only kind of prior knowledge that students bring to the learning situation. There is general domain knowledge that is not currently the object of study, but which can act as background knowledge for making sense of the domain and the inquiry question. There is also knowledge about the inquiry process that is not exclusively related to the domain, but which can equip students with appropriate methods for knowledge acquisition.

Purpose of the current study

The current study aimed to detect whether providing students with a partial hypothesis can facilitate their inquiry learning. To this end, the effectiveness of two versions of a hypothesis scratchpad as a support for students’ knowledge acquisition and inquiry process was compared. One version involved providing
students with terms representing the three main elements of a testable hypothesis (T); the other involved providing students with the same terms plus half a sentence giving the start of a hypothesis (T+PHy). Since a hypothesis plays an important role in guiding subsequent inquiry processes, including data collection and drawing conclusions, we examined the effect of the partial hypothesis not only on generating hypotheses, but also on the other inquiry processes. In addition, we were also interested to know if different types of prior knowledge influence the effect of the given hypothesis generation supports.

Specifically, the research questions examined in the study were:

1. What effect does providing students with partial hypotheses have on their hypothesis generation?
2. What effect does providing students with partial hypotheses have on their subsequent inquiry processes?
3. What effect does providing students with partial hypotheses have on their specific domain knowledge acquisition?
4. What influence do three different types of prior knowledge (knowledge about the inquiry process, general domain knowledge and specific domain knowledge) have on the effect of the given hypothesis generation supports?

Method

Participants

In total, 52 students from two secondary schools (20 students from School 1 and 32 students from School 2) completed all three sessions of the experiment. There were 25 boys and 27 girls, with a mean age of 13.87 years (SD = .99). To prevent a gender difference between the two experimental conditions, participants from each class were first grouped by gender and then randomly assigned to either experimental condition. This resulted in 23 students (11 male, 12 female) in the T (terms) condition, and the other 29 students (14 male, 15 female) in the T+PHy (terms + partial hypothesis) condition.

The experiment was carried out first at School 1, where it turned out that most students were not able to finish all of the inquiry phases. Hence, at School 2,
students were offered more time to work in the provided learning environment. The main arrangement and content of each session were identical for both schools; only the time allowed for each session differed. At School 1, each session lasted for 45 minutes, and at School 2, each session lasted for 60 minutes.

**Design**

A quasi-experimental pre-test/post-test design was used to examine the relative effectiveness of the two different versions (T; T+PHy) of the hypothesis scratchpad, a hypothesis generation support. The quality of students’ hypotheses generated, their performance on subsequent inquiry processes and their knowledge acquisition about the domain of force and motion addressed in the learning environment were the dependent variables, and the version of the hypothesis scratchpad was the independent variable.

**Materials**

In the following section, the supports, the learning environment and the measuring instruments used in present study will be introduced. All the materials were presented in English to the participants.

**Terms and partial hypotheses**

Terms and partial hypotheses were the two main supports for students’ hypothesis generation in this study. Students were provided with terms representing the three main aspects of a testable hypothesis: variables, relations, and conditions. The terms were based on the topic of Newton’s first law of motion and the current situation being simulated by a virtual lab on force and motion. For example, the terms provided for force variables were: the resultant force, the left force, and the right force, which included all types of forces that could be observed in the virtual lab. Based on these force variables, all relations that could occur between them were also offered as terms, such as larger, smaller, balanced and unbalanced. For the condition terms, the possible states of motion of an initially stationary object were provided: remains stationary and will move. Apart from the given terms, students were also offered an editable and reusable term box, in which students could type their own terms. They then needed to combine all of the chosen terms to form a complete hypothesis (Figure 2.1; Figure 2.2).
For the T+PHy condition, the first half-sentence of a hypothesis was provided. Based on the “If…then…” format of a hypothesis, the partial hypothesis in this study provided the “if” part of a hypothesis. This half-sentence clarified the condition of the independent variable, leaving students to complete this hypothesis by predicting the result of a relevant dependent variable (see an example in Figure 2.2). We provided the first half of the hypothesis rather than the predicted conclusion because it is logically more constrained to deduce a possible effect from a cause than the other way around. We assumed that it is easier for students to complete a hypothesis by stating a possible effect from the cause we provided in the first half-hypothesis, especially for those who do not know much about the domain being studied. The participants in the present study were assumed to have limited prior knowledge of the learning topic addressed in the present study, as this had not yet been taught in their physics course.

Figure 2.1. Screenshot of the T support in the hypothesis scratchpad in the Investigation 1 phase

Figure 2.2. Screenshot of the T+PHy support in the hypothesis scratchpad in the Investigation 1 phase
Providing partial hypotheses as support

Simulation-based learning environment

In this study, students worked in a simulation-based online learning environment. The inquiry learning environment was designed with an authoring platform named Go-Lab (Gillet, Rodríguez-Triana, de Jong, Bollen, & Dikke, 2017). An inquiry learning environment in Go-Lab is called an inquiry learning space (ILS); this term will be used in what follows. An introductory ILS and a main ILS were designed for both conditions. The introductory ILS was intended to familiarize students with the structure and the methods of operation of an ILS, while the main ILS was the main learning environment for the students. The introductory ILS had the same structure and the same form of hypothesis support as the main ILS, but addressed a different physics domain—electricity. Since the introductory ILS was designed mainly to offer practical guidance on how to operate in an ILS but not to give instruction on inquiry skills, students’ performance in the introductory ILS was not further analyzed.

In the main ILS, students were intended to learn about Newton’s first law of motion by working through 5 phases—Orientation, Preparation, Investigation 1, Investigation 2, and Reflection. The Orientation phase began by presenting an overview of the tasks to do in the ILS and providing basic knowledge about the domain. In a novel inquiry learning context, if the learning environment does not provide proper guidance and content information, even skillful and competent students will find difficulties in generating hypotheses. This phase aimed to prepare students with basic content information about the domain so that they could understand the research questions to be answered when writing their hypotheses. The information in this phase was mainly offered by text and pictures. A quiz with multiple-choice questions was included after each of several concept explanations. Both the correct answer and its explanation were provided as feedback for each of the answer options. These quizzes were presented to stimulate students to read the text and pictures carefully, and were not used for evaluation purposes. At the end of this phase, the learning goal of this ILS was clarified and students were prompted to proceed to the next phase.

The Preparation phase gave students background information on how to generate hypotheses and how to use the simulation—a Force and Motion lab (a PhET lab designed at the University of Colorado, retrieved from https://phet.colorado.edu/sims/html/forces-and-motion-basics/latest/forces-and-motion-basics_en.html). Multiple-choice questions were provided for students to review what they had learned about the concept and the format of a
testable hypothesis from the introductory ILS (3 items, e.g., What is a hypothesis? A. A hypothesis is a research question about what you are interested in; B. A hypothesis is a conclusion based on the results of an experiment; C. A hypothesis is a testable statement about your ideas concerning a research question; D. A hypothesis is an experiment designed to answer a research question). Students were given the correct answer and answer explanation after their responses. A summary of the main ideas concerning a hypothesis was given both in text and through a worked example of how to write a hypothesis. After this, an instructional video about how to use the Force and Motion Lab was shown to the student.

The Investigation 1 and Investigation 2 phases provided opportunities for students to examine the relationship between force and motion, with a focus on stationary objects and moving objects, respectively. In the Investigation 1 phase, students first watched an instructional video on how to use the hypothesis scratchpad. After that, they were asked to generate one hypothesis about the effect of force on the motion of a stationary object. The T or T+PHy support was provided to students in the hypothesis scratchpad (Figure 2.1; Figure 2.2). Then, in the Force and Motion lab, students could carry out experiments to test their hypotheses. In the Investigation 1 phase, students were asked to use the Net Force lab (a subordinate lab in the Force and Motion lab; Figure 2.3) to test their hypotheses about the effect of force on a stationary object. In this lab, students could apply force and see how the forces balance, how the resultant force keeps an object stationary or makes it move. The quantitative value and direction of the force were visible for students. While working with the labs, students could record their observed data in a data recording table (Figure 2.4). The aspects of information that should be recorded were specified in the first column of this table, to give examples of what data to record. After the data collection process, students were asked to write down their conclusion about whether to accept or reject their hypothesis in an input box presented right after the data recording table. In the Investigation 2 phase, students were asked to write a hypothesis about the effect of force on a moving object and to use the Motion lab (Figure 2.5) to test their hypothesis. This phase used almost the same structure as the Investigation 1 phase. Because specifying the aspects of information that should be recorded might limit students’ freedom in recording observations that were relevant to various hypotheses, the first column of the table was left open for the students in the Investigation 2 phase.
The last phase was the Reflection phase, where the students reflected on what they had learned about the relationship between force and motion. The conclusions written by the students at the end of the Investigation phases were (automatically) shown to them. Based on their findings from the previous phases, students were asked to draw their final conclusion about the relationship between force and motion. Prompts were also provided to motivate students to conclude in a direction that was in line with Newton's first law of motion.

The ultimate goal for supporting hypothesis generation is to facilitate the whole inquiry process. However, hypothesis generation is not the only challenging process for students in inquiry learning. Hence, apart from the hypothesis support, some supports for the other inquiry processes were also added in the ILS to guide students through the whole inquiry process, such as the data recording table in Investigation 1 phase and the prompts in the Reflection phase. One thing that should be noted here is that there could be an interaction effect between the hypothesis support and the other forms of support.

*Figure 2.3. Screenshot of the Net Force lab in the Investigation 1 phase*
Knowledge tests

Different tests were used to assess students’ different kinds of knowledge, including their prior knowledge about the inquiry process, their prior knowledge of the general domain (force and motion) and their knowledge about the specific topic (Newton’s first law of motion) covered in the learning environment. In the subsequent part, these three tests will be introduced one by one.
Providing partial hypotheses as support

**Inquiry process knowledge test**
In order to check students’ prior knowledge about the inquiry process, a paper and pencil test was used. This test included 8 multiple-choice questions (4 options), which were adapted from the Scientific Inquiry Literacy Test (ScInqLiT) by Wenning (2007). The test was checked by a physics teacher to make sure all of the questions were understandable for the students. One point could be earned for each question. The reliability of these 8 questions was .25 (Cronbach’s alpha).

**General domain knowledge test**
In order to check students’ prior knowledge of the general domain of force and motion, a paper and pencil test was used. General domain knowledge refers to a general level of domain knowledge that is related to the learning topic to be learned but does not address the learning topic per se. To be more specific, in this general domain knowledge test, 9 multiple-choice questions (4 options each) about force and motion, but not specifically about Newton’s first law of motion (the learning topic in this study), were included to check students’ general domain-related prior knowledge. These 9 questions were chosen and adapted from the Force Concept Inventory test originally designed by Hestenes et al. (1992) and a knowledge test on force and motion retrieved from https://www.warricksd.org/files/uploads/website/teacherweb/facultyfiles/alambert/Forces_and_Motion_Practice_Test.pdf. The questions chosen from the Force Concept Inventory were adapted into 4-option questions, rather than the original 5-option format. Several pictures were added in the test to illustrate the questions. Each item asked about a topic in force and motion, including gravity, friction, air resistance, action and reaction force, free-fall motion, uniform motion, relative motion, inertia and projectile motion. The reliability of these 9 questions was .52 (Cronbach’s alpha).

**Specific domain knowledge test**
To answer the first research question about whether providing students with a partial hypothesis can facilitate their knowledge acquisition, and to check students’ prior knowledge of the specific domain to be learned in the ILS, a specific domain knowledge test was designed. This knowledge test was also a paper and pencil test, which contained 14 multiple-choice questions with four options for each question. Both conceptual recall questions (4 items, e.g., A ________ is the sum of all of the forces acting on an object. A. balanced force; B. activated force; C. resultant force; D. direction of motion) and application questions (10 items, see Figure 2.6 for an example) were covered in this test. The content of this test was
also checked and commented on by the physics teacher. Modifications were made to make sure the questions were on the topic as well as understandable for students. This test was administered right before and after the intervention as a pre-test and post-test. The items on these two versions of the test were the same, but the orders of questions and options were changed. One point could be earned for each question. The reliability of the 14 questions was examined with Cronbach’s alpha. The reliability results for the pre-test and the post-test were .07 and .60 respectively. Specially, students’ performance on these three tests mentioned above were used to detect whether these three different types of prior knowledge impact the effect of the given hypothesis supports.

5. The following football shown is stationary (not moving) on the ground. Is there any force acting to this football now? _____

![Image of a football]

| A. Yes, there is one force acting to it | B. No, there is no force acting to it |
| C. Yes, there are two forces acting to it | D. None of the other options is correct |

6. Two people are pushing the same box in opposite directions; the amount of forces applied to the box is shown. Which of the following descriptions is right? _____

![Image of boxes with forces applied]

| A. A 15 N force acting to the right would have the same effect on the motion of the box | B. A 15 m per minute force acting to the right would have the same effect on the motion of the box |
| C. A 35 N force acting to the right would have the same effect on the motion of the box | D. A 35 m per minute force acting to the left would have the same effect on the motion of the box |

Figure 2.6. Example of the application questions on the domain knowledge test (correct answers are shown in bold)

The reliability of the three knowledge tests mentioned above was questionable, and the reliability of the specific domain knowledge pre-test was even extremely low. One reason that might account for this was that the tests did not measure one construct, but several sub-constructs within a general construct. For instance, the
Providing partial hypotheses as support

8 items in the test on knowledge of the inquiry process should all measure students' knowledge of the inquiry process. But inquiry process knowledge is a general construct that includes different sub-constructs, such as generating hypotheses, designing experiments, drawing conclusions, and so forth. The 8 items in this case actually test different sub-constructs of the same general construct, which might be the reason for the low interrelatedness of the items. Another reason might be that a small number of items were used in each test. The number of test items can affect the value of Cronbach’s alpha (Tavakol & Dennick, 2011). Because of the limited time allowed for the whole experiment, only a small number of items were included in each of the tests. Although quite low, the reliability, to some extent, matched the practical situation in this study.

**Coding and scoring**

To answer the first and second research questions on whether providing students with a partial hypothesis can facilitate their hypothesis generation and subsequent inquiry processes (data collection, drawing conclusions and final reflection), a coding scheme (see Appendix A) was designed to transform qualitative data on student behavior in the ILS into quantitative data. This scheme mainly covered the coding and scoring rules for four learning processes: hypothesis generation, data collection, drawing conclusions, and final reflection.

**Coding for hypothesis generation**

*Testability*

Four coding items were used to score the testability. Based on the topic of Newton’s first law of motion, as well as a tentative coding of one-third of the hypotheses, a list of possible variables and relevant conditions was developed. In the first half of a hypothesis, if a student mentioned one of the variables on the list, one point was assigned for the valid independent variable; if an observable or measurable condition of the variable was stated, another point was given for the valid condition of the independent variable. In the second half of the hypothesis, if a student mentioned one of the variables on the list and this variable was not the same as the independent variable, one point was allocated for the valid dependent variable; if an observable or measurable condition of the dependent variable was stated, another point was given for the valid condition of the dependent variable. For example, a hypothesis from the T condition could be: If the left force is larger than the right force, then the object moves to the left. In this case, the independent
variables are the left force and the right force. The condition of the independent variables is that the left force is larger than the right force. One point will be assigned for the valid independent variables and another point will be assigned for the valid condition. The dependent variable is the motion of the object. One point will be assigned for this valid dependent variable. Besides, since the hypothesis indicates the object’s direction of motion, another point will be given for the valid outcome condition. In particular, the comparability of the score coded from the generated hypotheses between conditions was also taken into consideration. Because the first half of the hypothesis was already given for the students in the T+PHy condition, the score on testability coded from the second half of the hypothesis was additionally compared between conditions.

**Complexity**

Two coding items were used to score the complexity of hypotheses. One item was about variable selection, concerning whether a relationship between force and motion was stated in the hypothesis rather than focusing only on either force or motion. Consider, for instance, hypothesis 1: If the left force is the same as the right force, then the sum of forces is 0; and hypothesis 2: If the left force is the same as the right force, then the object will stay stationary. In hypothesis 1, both the independent and dependent variables are about force, while in hypothesis 2, the independent variables are about force, and the dependent variable is about motion. The variable selection in hypothesis 2 will be regarded as more complex than that in hypothesis 1. Hypothesis 2 and hypothesis 1 will be assigned with one and zero point respectively for the variable selection. The second item was about the condition of the variable, concerning whether the hypothesis focused on a more generalized condition of force (balanced/unbalanced condition). Consider, for instance, hypothesis 1: If the left force is larger than the right force, then the object will move to the left; and hypothesis 2: If the left force and the right force are unbalanced, then the object will move in the direction of the larger force. The second hypothesis states a more generalized condition of the variables, which covers more specific situations. In this case, one point will be given to hypothesis 2.

**Coding for data collection**

Students’ data recorded in the data recording table were coded for three aspects. The main idea was to code whether the student actually recorded the data needed to test his or her hypothesis. One aspect was whether the condition of the independent variable mentioned in the hypothesis was recorded (e.g., the condition of the force); another aspect was whether the outcome condition of the
dependent variable was recorded (e.g., the outcome state of motion of the object); the last aspect concerned an important pre-condition that determines the correctness of the conclusion—the initial state of motion (stationary or moving) of the object. Each aspect accounted for one point. For instance, if the hypothesis is: If the left force is larger than the right force, then a stationary object will move to the left. If a student recorded the amount of force toward the left and right sides, and the left force is larger than the right force, then the student received one point. If the direction of motion of the object was recorded, the student got another point. In addition, if the student mentioned that object was initially stationary, then another point was awarded.

**Coding for drawing conclusions**

Students' conclusions were mainly coded for two aspects. One was whether a student’s conclusion was to accept a correct hypothesis or to reject an incorrect hypothesis. Another was whether the student’s conclusion could be inferred from the data he or she recorded. This second aspect was added to take a closer look at whether the student drew a data-based conclusion. A student could earn one point for each aspect.

**Coding for final reflection**

The reflection was coded based on the student’s final conclusion about the effect of force on motion in the Reflection phase. Two coding aspects were included, in line with the two main conclusions of Newton’s first law of motion: force can change motion and motion can be maintained by force. The students received one point if they generally mentioned each aspect. And the students could get two points if the condition in which force can change motion or in which motion can be maintained by force was correctly mentioned.

A second rater was trained to code the quality of students’ inquiry process based on the coding scheme, and coded the inquiry process of 12 students (23%). The interrater reliability coefficients for coding the inquiry process in terms of hypothesis generation in the Investigation 1 phase reached .91 and in the Investigation 2 phase reached 1.00 (Cohen’s kappa). The interrater reliability coefficient for coding the data collection in the Investigation 1 phase reached .72 (Cohen’s kappa) and in the Investigation 2 phase reached .75 (Cohen’s kappa). The interrater reliability coefficient was .71 and .88 (Cohen’s kappa), respectively, for drawing conclusions in the Investigation 1 and Investigation 2 phases. And the
interrater reliability for coding the final reflection in the Reflection phase was .81 (Cohen’s kappa).

Procedure

This study was conducted over 3 sessions on separate days. All of the participants followed the same sequence in the procedure (see Figure 2.7). The first session started with a brief introduction to the study, after which students were allowed 15 minutes to complete the knowledge of the inquiry process test and the general domain knowledge test. After the test, a short presentation took place to introduce some general operational skills for working with the ILS and guide students in logging into the introductory ILS. This introductory ILS had similar phases and hypothesis supports as the main ILS, but it was about another topic: electricity. Students were expected to get used to the main phases of inquiry learning as well as to practice their operational skills by exploring the learning environment in the remainder of this session. Students worked on their own computer and they were informed that they were free to ask questions if they came across any technical problems and that they should complete working in the ILS individually.

During the second session, students were first given 10 minutes to complete the specific domain knowledge pre-test. Then they were allowed to log into the main ILS about force and motion. During this session, only the first three phases (Orientation, Preparation, and Investigation 1) were visible for students. They were informed that they should go through the information carefully and complete the tasks in each phase one by one. In this session, they were allowed to ask questions about operational issues, but not about the learning domain.

In the last session, all five phases were visible for the students. They were first asked to take 5 minutes to look back at what they had done in the previous main ILS phases. Then they were allowed to start from where they had stopped last time and finish all of the phases in this session. The last 10 minutes were left for the specific domain knowledge post-test.

The physics teachers were present in the classroom to control the discipline of the classroom during all sessions. They were informed not to answer any questions about the learning domain or the inquiry task during the experiment. The experimenter was in charge of guiding the whole procedure and answering students’ questions about technical issues.
Providing partial hypotheses as support

![Diagram of experimental procedure](image)

Figure 2.7. An overview of the experimental procedure

## Results

### Impact of condition on knowledge acquisition

Table 2.1 summarizes the descriptive statistics for the main tests in this study.

<table>
<thead>
<tr>
<th>Data source</th>
<th>Maximum score</th>
<th>Condition</th>
<th>T</th>
<th>T+PHy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inquiry process knowledge test</td>
<td>8</td>
<td>T</td>
<td>4.70</td>
<td>4.86</td>
</tr>
<tr>
<td></td>
<td></td>
<td>T+PHy</td>
<td>4.86</td>
<td>4.72</td>
</tr>
<tr>
<td>General domain knowledge test</td>
<td>9</td>
<td>T</td>
<td>5.22</td>
<td>5.69</td>
</tr>
<tr>
<td></td>
<td></td>
<td>T+PHy</td>
<td>4.72</td>
<td>5.69</td>
</tr>
<tr>
<td>Specific domain knowledge test (pre-test)</td>
<td>14</td>
<td>T</td>
<td>6.17</td>
<td>7.22</td>
</tr>
<tr>
<td></td>
<td></td>
<td>T+PHy</td>
<td>5.69</td>
<td>8.34</td>
</tr>
<tr>
<td>Specific domain knowledge test (post-test)</td>
<td>14</td>
<td>T</td>
<td>7.22</td>
<td>7.22</td>
</tr>
<tr>
<td></td>
<td></td>
<td>T+PHy</td>
<td>8.34</td>
<td>8.34</td>
</tr>
</tbody>
</table>

The normality of students’ scores on the specific domain knowledge pre-test and post-test was checked. The results of the Shapiro-Wilk test were non-significant for both tests ($D(52) = .96, p = .08$ and $D(52) = .97, p = .13$, respectively), indicating that
the normality of scores on these two tests was confirmed. Therefore, parametric tests were used to analyze the results of the pre-test and post-test.

Students’ post-test score significantly increased from the pre-test in both the T+PHy condition \((t(29) = -6.21, p < .001, r = .76)\) and the T condition \((t(23) = -2.44, p = .02, r = .46)\), indicating that students in both conditions acquired domain knowledge from the ILS. The standard deviation of the post-test score also increased compared to that of the pre-test score (see Table 2.1), which means that the heterogeneity among students in specific domain knowledge level increased after the intervention. This finding suggests that the provided support did not benefit students equally; it impacted students’ knowledge acquisition differently for different students.

To examine whether condition influenced knowledge acquisition, an ANCOVA with the post-test score as the dependent variable and the specific domain knowledge pre-test score, the score on knowledge of the inquiry process and the score on general domain knowledge as the covariates was performed. Students’ specific domain knowledge \((t(50) = -1.12, p = .27)\), knowledge about the inquiry process \((t(50) = .41, p = .69)\), and their general domain knowledge \((t(50) = -.95, p = .35)\) were roughly the same between conditions, which showed that the assumption on the independence of the covariates was met. No significant interaction between conditions and specific domain knowledge \((F(1, 48) = .49, p = .49)\), knowledge about the inquiry process \((F(1, 48) = 3.26, p = .08)\) and general domain knowledge \((F(1, 48) = .65, p = .42)\), respectively, were found, indicating that the assumption about the homogeneity of regression slopes was met. The ANCOVA results revealed that there was a significant effect of different support conditions on students’ post-test score after controlling for the effect of specific topic knowledge, knowledge about the inquiry process and general domain knowledge \((F(1, 47) = 6.45, p = .01, \text{partial } \eta^2 = .12)\). Planned contrasts revealed that having the partial hypothesis significantly increased students’ mean post-test score compared to having only the terms \((t(47) = 2.54, p = .01, r = .35)\). These results indicated that providing students with T+PHy compared to only terms facilitated students’ acquisition of specific domain knowledge, with a medium effect size.

**Impact of condition on the inquiry process**

The quality of students’ inquiry processes was also compared between conditions. Table 2.2 shows the descriptive statistics for students’ scores coded from their hypothesis generation, data collection, and drawing of a conclusion in the
Providing partial hypotheses as support

Investigation 1 phase. Specially, to take a closer look at students’ hypothesis generation performance, as well as to make the coded score on hypothesis generation comparable between conditions, the testability and complexity of the full hypothesis and the score coded only from the second half hypothesis are both shown in Table 2.2. Because these scores were not normally distributed, Mann-Whitney tests were used to evaluate the differences in these processes between conditions.

It can be seen in Table 2.2 that the students did not differ much on their mean scores for the inquiry process between conditions. In terms of hypothesis generation, the results of Mann-Whitney tests revealed that students in the T+PHy group scored significantly higher on the complexity of the full hypothesis than students in the T group ($U = 222.00$, $z = -2.42$, $p = .02$), but non-significant differences were found on the testability of the second half of the hypothesis ($U = 307.00$, $z = -.82$, $p = .41$) and the full hypothesis ($U = 307.00$, $z = -.82$, $p = .41$). With regard to data collection ($U = 296.00$, $z = -.81$, $p = .42$) and drawing conclusions ($U = 305.50$, $z = -.64$, $p = .53$), no significant differences were found.

Table 2.2

Descriptive statistics for students’ coded scores for each inquiry process in the Investigation 1 phase

<table>
<thead>
<tr>
<th>Main variables</th>
<th>Maximum score</th>
<th>Condition</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>T (n = 23)</td>
<td>T+PHy (n = 29)</td>
<td></td>
</tr>
<tr>
<td>Hypothesis generation 1</td>
<td></td>
<td>M (SD)</td>
<td>M (SD)</td>
<td></td>
</tr>
<tr>
<td>Testability (second half of hypothesis)</td>
<td>2</td>
<td>1.83 (.58)</td>
<td>1.72 (.65)</td>
<td></td>
</tr>
<tr>
<td>Testability (full hypothesis)</td>
<td>4</td>
<td>3.83 (.58)</td>
<td>3.72 (.65)</td>
<td></td>
</tr>
<tr>
<td>Complexity (full hypothesis)</td>
<td>2</td>
<td>1.22 (.42)</td>
<td>1.55 (.51)</td>
<td></td>
</tr>
<tr>
<td>Data collection 1</td>
<td>3</td>
<td>2.30 (.97)</td>
<td>2.48 (.87)</td>
<td></td>
</tr>
<tr>
<td>Drawing conclusions 1</td>
<td>2</td>
<td>1.43 (.84)</td>
<td>1.59 (.73)</td>
<td></td>
</tr>
</tbody>
</table>

Since few students (6 out of 20) from School 1 managed to complete the inquiry process in Investigation 2, and nobody wrote the final reflection in the Reflection phase, students’ inquiry process in the Investigation 2 phase and their final reflection were analyzed separately for School 2. Table 2.3 shows the means and
standard deviations of students’ scores for the main inquiry process in the Investigation 2 phase. There was a trend that students in the T+PHy condition performed better on all of the listed process variables. Besides, students in the T+PHy condition earned all the points on hypothesis generation. The statistical results showed significant differences between conditions on the complexity of the full hypothesis \( (U = 8.5, z = -5.17, p < .001) \) and data collection \( (U = 63.50, z = -2.55, p = .01) \). No significant differences were found on the testability of the second half of the hypothesis \( (U = 110.50, z = -1.53, p = .13) \) and the full hypothesis \( (U = 110.50, z = -1.53, p = .13) \), on drawing conclusions \( (U = 87.50, z = -1.75, p = .08) \), and on the final reflection \( (U = 100.00, z = -1.08, p = .28) \).

Table 2.3
Descriptive statistics for the coded scores for each inquiry process in the Investigation 2 phase and the final reflection in the Reflection phase, for students from school 2

<table>
<thead>
<tr>
<th>Condition</th>
<th>Maxi-</th>
<th>M (SD)</th>
<th>M (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>T ( (n = 15) )</td>
<td>Hypothesis generation 2</td>
<td>2</td>
<td>1.80 (.56)</td>
</tr>
<tr>
<td>T+PHy ( (n = 17) )</td>
<td>Testability (second half of hypothesis)</td>
<td>4</td>
<td>3.67 (.90)</td>
</tr>
<tr>
<td></td>
<td>Complexity (full hypothesis)</td>
<td>2</td>
<td>1.00 (.38)</td>
</tr>
<tr>
<td></td>
<td>Data collection 2</td>
<td>3</td>
<td>1.27 (.70)</td>
</tr>
<tr>
<td></td>
<td>Drawing conclusions 2</td>
<td>2</td>
<td>.27 (.46)</td>
</tr>
<tr>
<td></td>
<td>Final reflection</td>
<td>4</td>
<td>1.33 (1.45)</td>
</tr>
</tbody>
</table>

In addition, since we assumed that hypotheses can be important references and guidance for the subsequent inquiry processes, the moderating effect of students’ score for hypothesis generation on the relationship between conditions of support and students’ scores for data collection and drawing conclusions were also examined. Since it was the probable cause and effect relationship conveyed by a
full hypothesis that may guide the data collection and drawing conclusions processes, students’ score coded from the full hypothesis (sum of score on testability and complexity) was examined as the moderator variable.

Linear regression based on the data from the Investigation 1 phase showed a significant interaction effect of condition and hypothesis generation on students’ data collection score ($\beta = 2.39, SE = .27, t = 3.01, p = .00$), but a non-significant interaction effect on students’ score for drawing conclusions ($\beta = 1.28, SE = .26, t = 1.42, p = .16$). These results indicated that there was a significant moderation by hypothesis generation of the effect of condition on students’ data collection.

The same method was used on the data from the Investigation 2 phase, which yielded a non-significant interaction effect of condition and hypothesis generation on students’ data collection ($\beta = -.21, SE = .06, t = -1.31, p = .20$) and drawing conclusions ($\beta = .09, SE = .18, t = .38, p = .71$).

**Impact of prior knowledge**

To check whether students’ prior knowledge influenced the effect of the different versions of hypothesis support, the moderating effect of the three types of prior knowledge (knowledge about the inquiry process, general domain knowledge, and specific domain knowledge) on the strength of the relationship between the predictor variable (condition) and the dependent variables (specific domain knowledge post-test score and total score for the inquiry process) were examined. Table 2.4 shows the descriptive statistics for students’ total score for the inquiry process in each Investigation phase, which was the sum of students’ scores for the complexity and testability of hypotheses, data collection and drawing conclusions.

Table 2.5 presents the correlations (Pearson correlation) between the different test results. The correlational results indicate that students’ knowledge about the inquiry process and their specific domain knowledge before the intervention significantly correlated with students’ specific domain knowledge after the intervention; there was no significant correlation between general domain knowledge before the intervention and specific domain knowledge at post-test.
Table 2.4
Descriptive statistics for students’ total score for the inquiry process in the Investigation 1 and Investigation 2 phases

<table>
<thead>
<tr>
<th>Condition</th>
<th>T (n = 23)</th>
<th>T+PHy (n = 29)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Main variables</strong></td>
<td><strong>Maximum score</strong></td>
<td><strong>M (SD)</strong></td>
</tr>
<tr>
<td>Inquiry process 1</td>
<td>11</td>
<td>8.78 (2.26)</td>
</tr>
<tr>
<td>Inquiry process 2</td>
<td>11</td>
<td>6.20 (1.74)</td>
</tr>
</tbody>
</table>

Table 2.5
Correlations between the knowledge test scores (Pearson correlation)

<table>
<thead>
<tr>
<th>Variables</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Knowledge about the inquiry process</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. General domain knowledge</td>
<td>.18</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Specific domain knowledge (pre-test)</td>
<td>.03</td>
<td>.17</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>4. Specific domain knowledge (post-test)</td>
<td>.34*</td>
<td>.26</td>
<td>.36**</td>
<td>1</td>
</tr>
</tbody>
</table>

Note. ** = p < .01, * = p < .05. N = 52 for all knowledge tests

The results of a simple linear regression between conditions and the specific domain knowledge post-test score indicated that condition was not a significant predictor of the post-test score ($\beta = -.24, SE = .65, t = -1.74, p = .09$). When adding students’ knowledge about the inquiry process as a second predictor, and the interaction between condition and knowledge about the inquiry process as the third predictor, the interaction variable was also a non-significant predictor of the post-test score ($\beta = -1.01, SE = .42, t = -1.81, p = .08$). Similar results were found for students’ general domain knowledge ($\beta = .47, SE = .36, t = .81, p = .42$) and specific domain knowledge ($\beta = -.48, SE = .39, t = -.70, p = .49$). There was also no moderating effect on the relationship between condition and students’ post-test score.

The same tests were used to check prior knowledge as a moderator of the relationship between condition and students’ total score for the inquiry process in each Investigation phase. No significant moderating effect of students’ knowledge about the inquiry process ($\beta = -.45, SE = .39, t = -.76, p = .45$), general domain
Providing partial hypotheses as support

knowledge ($\beta = .83, SE = .31, t = 1.45, p = .15$), and specific domain knowledge ($\beta = .25, SE = .39, t = .32, p = .75$) was found on the relationship between condition and students’ inquiry process scores in the Investigation 1 phase. The results for the Investigation 2 phase were also non-significant ($\beta = -1.16, SE = .55, t = -1.58, p = .13; \beta = .63, SE = .34, t = 1.02, p = .32; \beta = .76, SE = .47, t = .86, p = .40$, respectively).

**Discussion**

The present study investigated whether providing students with a partial hypothesis is an effective way of promoting their knowledge acquisition, the quality of their hypotheses, and their performance on other subsequent inquiry processes, over and above providing them with a set of terms in the hypothesis scratchpad. The impact of three types of prior knowledge on the effect of the given partial hypothesis on knowledge acquisition and the inquiry process were also examined. The participants either received as support a set of terms representing the three main elements (variables, conditions, and relations) that could be used to write hypotheses (T), or received an identical set of terms plus the start of a partial sentence stating a hypothesis (T+PHy). We assumed that students who worked with T+PHy would outperform students who worked only with T on knowledge acquisition, hypothesis generation, data collection, drawing conclusion and final reflection.

Although the reliability of the knowledge tests is worth further consideration, differences regarding knowledge acquisition were found between conditions as expected, suggesting that providing students with partial hypotheses indeed fostered students’ knowledge acquisition. This is in line with the previous findings that providing partial supports can facilitate students’ learning performance (Chang et al., 2001; Mulder et al., 2016). One possible explanation may be that the presented partial hypothesis potentially directed students’ attention to more important and more productive parts of the learning task, which may have deepened students’ understanding of the domain. This suggestion is supported by the difference found between conditions on the quality of hypotheses.

To be more specific, a difference was found in the complexity but not in the testability of the hypothesis between conditions. In the T condition, students could write down their first tentative ideas about force and motion as their hypothesis, which can be quite simple and testable. This can be inferred from the fact that almost half of the students wrote very simple, similar hypotheses, such as, “If the
left force is larger than the right force, the stationary object will move to the left”, which was probably too easy as a starting point for the inquiry process. In contrast, in the T+PHy condition, students first needed to make sense of the condition provided for the independent variable in the first partial hypothesis, and then predict a possible outcome condition of a relevant dependent variable in the domain, which can be more challenging but also more meaningful than writing a hypothesis from scratch. In this study, the provided partial hypothesis in both Investigation phases specified a balanced state of forces, forcing students to consider the state of motion of an object when the applied forces are all balanced. A completed hypothesis example from the T+PHy condition was “If the resultant force on a stationary object is 0, then the stationary object remains stationary”. Hence, it is not surprising that students differed on the complexity but not on the testability of hypotheses. These findings may also account for why providing partial hypotheses assists students to have better knowledge acquisition. With the help of the partial hypothesis, students were likely to start their learning from a more meaningful point, thus promoting their knowledge acquisition. In this study, students were allowed limited chances to write hypotheses, which may have hindered them from exploring the specific domain more thoroughly. In future research, if we offer more partial hypotheses for students to complete, we can also fade the hypothesis support and detect the transfer effect of the given support.

The hypothesis that providing students with a partial hypothesis can further facilitate students’ subsequent inquiry processes was partially supported by the results. Only the score on data collection in the Investigation 2 phase was found to differ significantly between conditions. In the Investigation 1 phase, no significant differences were found. An explanation could be that the data recording table provided in the Investigation 1 phase compensated for the missing support in the T condition. In the Investigation 1 phase, prompts were provided in the first column of the data recording table to offer an example of what data to record. In the Investigation 2 phase, though, students were asked to decide what to record in a blank data recording table, having seen the examples in the table from the Investigation 1 phase. Students’ data collection performance differed between conditions when there was a blank data recording table, but not with the guided data recording table, suggesting that the inquiry direction offered by the given partial hypothesis can offer guidance on what to record for data collection. It should be noted that since only a few students from school 1 finished the Investigation 2 phase, the analysis of the inquiry process in Investigation 2 phase was based on the
Providing partial hypotheses as support

participants from school 2 only. Further research with a larger sample size should be implemented to examine the generalizability of the conclusion.

The impact of prior knowledge on the effect of the experimental conditions was also checked, but no significant moderating effect of any of the three types of prior knowledge (knowledge about the inquiry process, general domain knowledge, and specific domain knowledge) was found.

Regarding students' knowledge about the inquiry process, a potential explanation is that the extra process supports in the ILS might narrow the effect of prior knowledge of the inquiry process between conditions. Apart from the hypothesis support, students were also provided with prompts before working with the data recording table to guide students to record their data, and before responding to the conclusion input box to guide them to write their conclusions. The extra support in the current study might have compensated for students' inadequate knowledge about inquiry, especially those with lower knowledge, which might to some extent attenuate the effect of the level of knowledge about the inquiry process assessed at the very beginning.

As for general domain knowledge, one explanation might be that the general domain knowledge tested in the knowledge test was not the very important knowledge needed for students to make sense of the specific domain in this study. The fundamental goal of inquiry learning is that students need to develop knowledge of an unfamiliar task or domain (Lazonder, Wilhelm, & van Lieburg, 2009). Students can construct scientific conceptions if the new domain first makes sense to the students, and then encourages students to question their own conceptions and build their own perspectives. Hence, it was assumed that students with less general domain knowledge about force and motion (such as gravity, friction, free fall motion) might be struggling more with the specific domain that concerning the relationship between force and motion, so there is more chance that they might benefit from the direction offered by the given partial hypothesis. Yet, the results indicate that this was not the case. The measured general domain knowledge might not be the prerequisite knowledge that would influence students' need for extra hypothesis support. In the future, it may be helpful to specify and reliably assess the type or level of prior knowledge that can facilitate students' inquiry learning.

Concerning students' specific domain knowledge, the domain knowledge provided in the ILS might account for the nonsignificant results. The development
of domain knowledge and scientific reasoning skills requires students to have at least a basic understanding of the inquiry domain (Lazonder, Hagemans, & de Jong, 2010). Hence, in the Orientation phase of the ILS, students were provided with some domain-specific knowledge, aiming to prepare students with some background information to understand the question to be inquired about. Yet, students’ specific domain knowledge was tested before students started to work in the main ILS. There is a chance that students’ prior knowledge level changed after the Orientation phase, which would not be reflected in the pre-test used in the present study. In future research, more attention should be paid to the potential influence of the given domain knowledge provided in the ILS on the level of prior knowledge being tested.

To sum up, with the support of the given half-hypotheses, students can generate more complex hypotheses, perform better in data collection, and acquire more specific domain knowledge. The present study helps to advance educational design science one step further and offers a concrete suggestion on how to support students’ hypothesis generation. Apart from this, this study successfully applied the completion strategy promoted by prior researchers in a new situation, which also contributes to testing the generalizability of this support strategy.

References


Providing partial hypotheses as support


Chapter 2


Chapter 2
Chapter 3
Providing domain information and/or self-exploration as support

Abstract

This study investigated the effects of presenting domain information either together with or instead of offering exploratory practice for facilitating students’ hypothesis generation and subsequent inquiry processes and their knowledge acquisition. Secondary school students ($n = 118$) completed an inquiry task on force and motion. They were randomly assigned to one of four conditions: the D+E condition ($n = 29$), in which domain information and exploratory practice were available; the D condition ($n = 30$), in which only domain information was available; the E condition ($n = 32$), in which only exploratory practice was available; or the C condition ($n = 27$), with no support at all. Results indicated that providing students with domain information alone helped them to know more about the variables in the domain before generating hypotheses than those who received no additional support. There were no differences between conditions on the quality of hypothesis generation, subsequent inquiry processes, or knowledge acquisition.
Chapter 3

**Introduction**

Recent years have seen a growing call for increasing students’ involvement during science learning (Bond & Bedenlier, 2019; de Jong, 2019; Lei et al., 2018). Inquiry learning, a learning method in which students infer knowledge about a domain by formulating hypotheses, designing experiments, evaluating evidence and drawing conclusions (Lazonder et al., 2010; van Joolingen & de Jong, 1991), actively involves students in the learning process. Although inquiry learning is generally assumed to foster deep and meaningful science understanding (de Jong, 2006; Goodyear et al., 1991; Hmelo-Silver et al., 2007; Mayer, 2004; Reid et al., 2003), the effectiveness of this learning method has been hampered by students’ lack of the cognitive and practical abilities required to carry out this multifaceted learning activity. Students often have difficulties with completing inquiry processes and engaging in fruitful inquiry learning (J. Chen et al., 2018; de Jong & van Joolingen, 1998; Ferrés Gurt & Marbà Tallada, 2018). Multiple meta-analyses on the effectiveness of inquiry learning have concluded that students need support to benefit from inquiry learning (Alfieri et al., 2011; Belland et al., 2017; D’Angelo et al., 2014; Furtak et al., 2012; Lazonder & Harmsen, 2016; Minner et al., 2010; Rutten et al., 2012).

This study focuses on support for one crucial element of inquiry learning: hypothesis generation. In the present study, a hypothesis refers to a tentative and testable explanation of a phenomenon that states an empirical relation among variables in a given problem context (Quinn & George, 1975). Hypothesis generation is considered to be one of the key processes of inquiry learning (Klahr & Simon, 1999; Njoo & de Jong, 1989). Generating hypotheses triggers students to consider a problem seriously and helps them to organize ideas about their tentative answers or solutions for the given problem. The hypotheses generated then direct the designing of experiments and can act as standards in checking the validity of collected data (Wenham, 1993). When drawing conclusions about the findings in inquiry learning, the hypotheses can act as plausible forecasts with which to contrast the gathered evidence (Guisasola et al., 2006). In this way, the hypothesis generation process plays a fundamental role in guiding the subsequent processes of inquiry learning. That is why inquiry learning can also be described as a hypothesis-generating and hypothesis-testing procedure (Zimmerman, 2007). Yet, hypothesis generation is problematic for students, as they find it hard to formulate informative and testable hypotheses (de Jong & van Joolingen, 1998; Gijlers & de Jong, 2009). Students with low prior knowledge regarding a given
problem will fail to identify relevant variables, and they find it hard to generate initial hypotheses (Kim & Pedersen, 2011; van Joolingen & de Jong, 1991). It is also difficult for students to indicate potential relations between variables (de Jong & van Joolingen, 1998).

To determine the appropriate type of support for students’ hypothesis generation process, it is important to know the prerequisites for generating relevant and useful hypotheses. The basic principle of inquiry learning is to put students in a position to infer knowledge about a task or domain with which they are relatively unfamiliar (Lazonder et al., 2009). Inquiry learning yields knowledge, but for inquiry learning to function properly, a certain level of basic knowledge about the task or domain being explored is needed (Alexander & Judy, 1988; Hattie & Donoghue, 2016; Hulshof & de Jong, 2006; Lazonder et al., 2008; Sandoval & Reiser, 2004; Schneider & Preckel, 2017; Wallace et al., 2003). The knowledge that each student brings to a given learning context can serve as an interpretive framework that they can use to judge what variables need to be considered, to explain new observations (Jones et al., 2000; Sandoval & Reiser, 2004; Tabak et al., 1996), or to situate and anchor new knowledge (Novak, 2002). Students with little or even no prior basic knowledge of a domain lack the necessary knowledge framework to construct further understandings about the domain.

The SDDS (scientific discovery as dual search) model of Klahr and Dunbar (1988) indicates two sources of knowledge that can be the basis for construction of hypotheses. These authors argued that students’ inquiry learning activity is a search process in two related spaces: the hypothesis space and the experiment space. The hypothesis space contains all possible hypotheses about the domain under study, and the experiment space consists of all experiments that can be carried out within the system being investigated. According to this model, students begin their inquiry learning process from one of these two spaces. They can either start by generating hypotheses based on prior knowledge and then use experimental evidence to test their hypotheses, or start by first doing some exploratory experiments and then use what they have observed to guide their hypothesis generation. The two knowledge sources for hypothesis generation described in this model suggest two ways to support students’ hypothesis generation: one is to provide basic domain knowledge for students, and the other is to offer students the opportunity to do exploratory experiments. The main purpose of this study was to investigate whether providing these two sources of
knowledge together or providing either of them separately can facilitate students’ performance in hypothesis generation.

**Theoretical background**

A meta-synthesis by Hattie and Donoghue (2016) concluded that when students have not acquired the surface-level knowledge on which it is important to base deep conceptual knowledge, involving students in more engaged forms of learning (including inquiry learning) shows less favorable effect sizes. Another meta-synthesis by Schneider and Preckel (2017) found that prior instruction is a moderator variable that can influence the effectiveness of more engaged forms of learning such as inquiry learning. Furthermore, the overall results of the PISA 2015 survey (OECD, 2016) showed a negative correlation between the frequency of inquiry learning activities and students’ STEM performance. However, further, more detailed analyses found that there was not a linear, but a curvilinear relation between the frequency of inquiry activities and STEM performance; the frequency of inquiry activities was positively related with STEM performance up until a certain level (L. Chen et al., 2017). This finding indicates the importance of providing more direct forms of instruction together with inquiry learning; there should be room for preliminary direct instruction along with inquiry learning. From this body of work, it becomes clear that students need to possess some knowledge relevant to the domain before they can benefit from inquiry learning in a way that will facilitate further domain knowledge acquisition.

Although domain knowledge is critical for hypothesis generation, not every student possesses enough domain knowledge before starting inquiry learning. A number of studies have therefore investigated the effectiveness of presenting domain information for facilitating inquiry learning. Lazonder et al. (2010) provided domain help-files and a self-study guide on the variables in the inquiry tasks either before, or before and during, or not at all during the inquiry learning process. They found that students who were provided with information about variables relevant to the domain generated more specific hypotheses and gained more knowledge from the inquiry learning task than students without access to the domain-related information. Wecker et al. (2013) provided a detailed pre-specified script together with a teacher-led discussion of important variables for a learning topic prior to the inquiry learning task, and reported positive effects on students’ development of deeper theoretical understanding of the domain. A study by Reid et al. (2003) and a follow-up study by Zhang et al. (2004) both found
that access to and activation of relevant domain knowledge helped students to construct deeper and more elaborate understanding of the explored domain. In a study by Kukkonen et al. (2014), a preparatory lesson was designed to provide students with some background knowledge of the variables to be considered in an inquiry learning task. The results indicated that ensuring an adequate level of background information (i.e., information about variables involved) before the inquiry session enriched the concepts students used in their representations, leading to better understanding of the learning domain.

What kind of domain information needs to be provided? A study by Lazonder et al. (2008) investigated how students’ domain knowledge impacts the type of investigative strategy they use in inquiry learning within that domain. They compared students’ strategy use when performing a concrete inquiry task in a familiar domain and an abstract inquiry task for which they had no intuitive prior domain knowledge. The results indicated that students performed the concrete task by generating hypotheses first and then used experiments to test them, while they performed the abstract task by inferring knowledge from doing exploratory experiments first and then formulating their hypotheses. Students’ performance scores for detecting the relations between variables were higher in the concrete inquiry task than in the abstract inquiry task. A follow-up study by Lazonder et al. (2009) further found that it is not only the meaning of the variables per se that matters in facilitating hypothesis generation, but the understanding of the way these variables are interrelated. Similarly, in the problem-solving domain, Beckmann and Goode (2014) also found that a concrete problem context can foster students’ hypothesis generation performance. The main problem for students in their study to solve was to learn about the causal structure of a linear system with three input variables and three output variables. The problem was presented to students with different levels of semantic familiarity, through the use of different variable labels. Students’ sense of familiarity with the variables was based on their prior domain knowledge of the semantic labels. The results indicated that a semantically familiar problem context can foster students’ hypothesis generation performance, while lack of familiarity with the context can impede knowledge acquisition. These findings imply that declarative knowledge about the meaning of variables in a domain is important for successful inquiry, but that some ideas about the relations between these variables may also play a role. The two interventions that we used in this study each have a different emphasis. When providing students with direct information, the emphasis is on the knowledge of
variables. When giving students the opportunity to explore, there is more emphasis on the relations between variables.

Offering exploratory practice may be a way to bridge students’ knowledge gap in the relations between variables when generating hypotheses. There is a strong consensus that students understand scientific concepts more deeply when they encounter and use the concepts in a personally meaningful context (Dalgarno et al., 2014; DeCaro & Rittle-Johnson, 2012; Resnick, 1998). A period of exploratory practice enables students to activate their prior knowledge about the task domain and triggers them to encode new information in a more meaningful and relevant way (DeCaro & Rittle-Johnson, 2012). Allowing them to explore a learning environment in a target domain may make students familiar with the variables involved and prepare them with some basic ideas on the possible relations between variables in the domain, which may further benefit their inquiry process. With traditional hands-on physical materials, exploratory practice with concepts or variables can be relatively expensive and time-consuming (van Joolingen & de Jong, 1991). Computer simulations provide simplified and/or enhanced representations of the variables, which makes exploratory practice more convenient and accessible for students (van Joolingen et al., 2007). In addition, simulations can visualize variables or processes that are scarcely possible in physical experiments (e.g., friction-free ground), which allows students to manipulate a wider range of scientific phenomena (Windschitl, 2000). Hence, providing students with opportunities to explore a simulation prior to inquiry learning can be a promising way to promote their inquiry performance.

**Research goal and hypotheses**

Many empirical studies of inquiry learning have tended to focus on the support provided during inquiry processes (J. Chen et al., 2018; Kim & Pedersen, 2011; Mulder et al., 2016; Njoo & de Jong, 1993; Windschitl, 2000), while few studies have highlighted the need for instructional support prior to the inquiry learning. Presenting domain information has proved to be an effective way to facilitate inquiry learning (Lazonder et al., 2010), but no studies have investigated the combined effect of providing both domain information and exploratory practice prior to the inquiry learning process.

The major purpose of this study was to detect whether providing domain information together with exploratory practice can benefit students more than
Providing either of them alone, as far as improving their hypothesis generation performance and knowledge acquisition in a simulation-based inquiry learning environment. As already mentioned, inquiry learning is a process of hypothesis generation and hypothesis testing (Zimmerman, 2007). Hypotheses play an important role in guiding subsequent hypothesis testing processes, such as data recording and drawing conclusions. Hence, we were also interested in detecting the impact on subsequent inquiry processes of providing domain information together with or instead of exploratory practice. Toward this end, this study compared students’ inquiry processes and knowledge acquisition across four conditions. In the domain information plus exploratory practice (D+E) condition, both the meaning and examples of all of the relevant variables and an opportunity to explore the simulation lab were provided prior to inquiry learning. In the domain information only (D) condition, only the domain information was provided prior to inquiry learning. In the exploratory practice (E) condition, students were allowed to explore the simulation lab prior to inquiry learning, without any domain information being provided. The final condition was a control (C) condition, in which students were provided with neither domain information nor the opportunity for exploration prior to beginning their inquiry learning task.

The presence of domain information was expected to enlarge students’ knowledge base related to the domain and subsequently providing exploratory practice aimed to consolidate their understanding of the variables in the domain and to offer students a chance to get an idea of possible relations that could exist between variables. The knowledge thus acquired about variables and general ideas of how variables could be interrelated could be further used to generate hypotheses. Given that knowledge about the meaning of variables and knowledge about the possibilities for how variables are interrelated are both crucial to successful inquiry learning and hypothesis generation (Alexander & Judy, 1988; Lazonder et al., 2009; Sandoval & Reiser, 2004), students in the D+E condition were expected to generate a higher number of hypotheses, have a higher diversity of variables in their hypotheses, and generate more informative hypotheses than students in the other conditions. Students’ hypotheses, in turn, were expected to have an effect on their subsequent inquiry processes, which would further lead to better knowledge acquisition than in the other conditions. Moreover, it would make more sense to present domain information to students who started out with less prior basic knowledge about the learning topic. In the present study, we chose a learning topic that had not yet been taught to students. Cognitive load theory (Sweller et al., 1998) suggests that exploratory practice may cause heavy cognitive load that is
detrimental to learning for novice students, who lack proper schemas to integrate the new information into their prior knowledge. Students in the E condition were assumed to lack prior schemas to understand new variables and, according to cognitive load theory, would find it hard to infer possible relations between variables from their exploratory practice, whereas the D condition would prepare students with knowledge about variables to build their hypotheses on. Hence, although we expected students in the D condition to perform less well than those in the D+E condition, we assumed that they would outperform those in the E condition on hypothesis generation, other subsequent inquiry processes, and, finally, knowledge acquisition. Moreover, all experimental conditions were expected to lead to better results than the control condition.

Method

Participants

Participants were 144 students, between 13 and 14 years old, from eight classes of grade 9 (key stage 3) at a secondary school in the UK. All participants returned parental consent forms. Twenty-two students who missed one or more sessions of the experiment as well as four other students who wrote one hypothesis or even no hypotheses without going on to test them were removed from the data analysis. The final sample consisted of $N = 118$ participants (mean age = 13.54 years, $SD = .50$, 61.9% male). Participants from each class were randomly assigned to either the domain introduction + exploration (D+E) condition ($n = 29$), the domain introduction only (D) condition ($n = 30$), the exploration only (E) condition ($n = 32$), or the control (C) condition ($n = 27$).

Learning environment

Students worked through a sequence of learning phases in a simulation-based inquiry learning environment on a physics topic—Newton’s first law of motion. This topic was chosen together with the physics teachers of the target school. The topic was part of the regular curriculum that had not yet been taught to the students and that would normally be taught by explicit instruction. The learning environment was designed with the Go-Lab ecosystem (de Jong et al., 2021), which provides access to a large collection of simulation labs, scaffolding apps, and tools for inquiry learning. Students had no prior experience with working in this type
Providing domain information and/or self-exploration as support of inquiry learning environment. The learning environment used in the present study included six learning phases that were largely aligned with the inquiry cycle introduced by Pedaste et al. (2015): (1) Introduction, (2) Domain information (What is force?; What is friction?; and What is motion?), (3) Exploration, (4) Hypothesis generation, (5) Investigation, and (6) Conclusions. The first three phases were not provided for all of the students. As shown in Table 3.1, in line with the intended treatments for the four conditions, the initial learning phases provided for students differed across conditions.

Table 3.1
The availability of the phases in the learning environment for each condition

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>What is force?</td>
<td>What is friction?</td>
<td>What is motion?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D+E</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>D</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>E</td>
<td>√</td>
<td></td>
<td>√</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>C</td>
<td>√</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Introduction**

The Introduction phase presented the learning topic and learning goals to students briefly with a picture and text. The picture showed different daily situations involving force and motion. Several questions about force and motion (e.g., Do you know what force is? What motion is? Do you know how forces affect the motion of an object?) were provided to introduce the target learning topic. No further information about variables for force and motion were presented in this phase. This phase aimed to stimulate students to think about force and motion and to activate their prior knowledge about force and motion.

**Domain information**

The Domain information phase included three sub-phases, which were titled: What is force?, What is friction?, and What is motion? These three sub-phases introduced definitions of the three main concepts in the domain, as well as the variables related to those concepts. For instance, in the What is force? phase, the definition of force was stated in words and a picture was used to illustrate the
definition. A few specifics about the two variables related to force (e.g., the unit of magnitude of a force, and the direction of the force) were explained by a picture example and a corresponding explanation (see Figure 3.1). Within these three sub-phases, a total of six variables in the domain that could be observed or manipulated in the learning environment were introduced, namely: magnitude of a force, direction in which a force is applied, friction, speed of a moving object, direction of motion of an object, and the initial state of motion. The relations among variables were left undisclosed for students and had to be explored by generating and testing their hypotheses in later phases. This phase aimed at ensuring that students possessed enough knowledge about variables to generate good hypotheses.

Figure 3.1. Screenshot of one of the domain information phases – What is force?

**Exploration**

In the Exploration phase, students were first shown how to use the simulation lab on force and motion (a simulation lab designed by PhET Interactive Simulations, University of Colorado Boulder, licensed under CC-BY 4.0. https://phet.colorado.edu.) by using guiding pictures (see an example in Figure 3.2). The use of guiding pictures in this phase instead of video guidance was done intentionally, in order to avoid giving away the relations between variables when applying forces and showing the outcome state of motion to students, as would be the case in an instructional video. After the picture-based instructions, students were given access to the simulation lab and were allowed to work with the lab for 10 minutes. Students were asked to get familiar with how the lab works and to adjust
variables to see how the variables related to force and motion are simulated in the lab and to get a basic idea of how variables were related.

Figure 3.2. Screenshot of an example of picture-based guidance on how the force and motion lab works

**Hypothesis generation**

The Hypothesis generation phase aimed to guide students in writing their hypotheses on how forces affect the motion of an object. A worked-out example with a different topic (plant growing) was provided to illustrate what a testable hypothesis looks like. An instructional video on how to use the Hypothesis Scratchpad, a hypothesis generation tool initially designed by van Joolingen and de Jong (1991), followed. In the Hypothesis Scratchpad, the three main elements of a testable hypothesis (variables, relations and conditions) were provided as terms with which students could build their hypotheses. The variable terms covered the six major variables introduced in the Domain information phase, while the condition and relation terms covered the basic conditions and relations that could occur in or between the variables. Apart from the given terms, an editable and reusable term space was also provided for students to express their own ideas (see Figure 3.3). The hypothesis scratchpad does not offer any information on the meaning of the variables and the relations offered in the scratchpad had a general nature. In all cases, students still have to choose to use the variables and relations themselves. Students were asked to write at least five hypotheses in the five given
boxes, and they were informed that more hypotheses could be included by clicking the “+” button in the tool.

Figure 3.3. Screenshot of the Hypothesis Scratchpad with an example hypothesis

**Investigation**

The Investigation phase was for students to test their hypotheses. Students were first shown the hypotheses that they had created in the previous phase. Below these hypotheses, hints on experiment design were provided to guide students on how to use the virtual lab to test their hypotheses (e.g., when you design an experiment, change one thing at a time; use the pause button to make it easier for you to record what you observed from the lab; reset the lab every time you start a new experiment). Then, hints on how to record their data were offered (e.g., record data on what you observed from the force and motion lab to test each of your hypotheses; when you want to record a change of motion, please record both the initial state of motion and the following state of motion to show the change) and an observation tool was included for students to record their data. This tool enabled students to insert input boxes, in which they could type what they observed from the virtual lab. Below the observation tool, the virtual lab was provided for students to work with. The hints and the observation tool were provided before showing the virtual lab, which was based on the consideration that hints on experiment design and data recording could inform students about the learning goals for what they were supposed to do in the virtual lab.
Conclusions

Students’ generated hypotheses and the recorded data were automatically shown at the beginning of this phase. Students were asked to draw conclusions on whether to accept or reject their hypotheses, one by one. At the end, students were asked to summarize how forces affect the motion of an object by answering two open-ended questions.

Procedure

The experiment was conducted over three 55-minute sessions held at the regular time for physics lessons. The sessions were led by three trained experimenters. All of the experimenters followed the same lesson plan and each class was led by the same experimenter for all the three sessions.

Prior to the first session, a roundtable meeting of all of the physics teachers involved and the experimenters took place to inform the teachers about the basic purpose of the experiment and about their roles during the experiment. The physics teachers were instructed not to answer content-related questions from their students, and were told that students were not allowed to leave if they finished the inquiry tasks early. The teachers were present during the three sessions to maintain classroom discipline.

At the beginning of the first session, students were randomly assigned to the four conditions by being given a group number when entering the classroom. The paper indicating the group number also gave the link to the learning environment and a login code. Students sat in their assigned groups and each of them sat at their own computer. After a brief introduction to the major experimental procedures, students were allowed 8 minutes to complete the general domain knowledge pre-test (see the Measures section). Then an example of an inquiry learning environment on electricity was shown to them phase by phase to demonstrate the layout and the basic operational skills used in the learning environment. Then students logged into the learning environment and started the inquiry learning task. Since different conditions differed in the number of phases to complete before students could start on their inquiry processes, the time for administration of the variable-related knowledge test (see the Measures section) also differed among conditions. Students were given a reminder at the appropriate point (at the end of the Introduction for those in the C condition, after seeing the Domain information for those in the D condition, and after Exploration for those in the E and D+E conditions) to raise their
hands to notify the experimenter that they had finished reading the information. The students were asked to log out of the learning environment before they started to complete the test and they were given more time if they indicated that they needed it. After handing in the variable-related knowledge test, students were told to continue working on the inquiry learning task.

The second and third sessions followed almost the same procedure. Students were first asked to review the previous phases and then start from where they stopped in the last session to continue working on the inquiry learning task. The only difference was that the last 8 minutes of the third session were for students to complete the general domain knowledge post-test (see the Measures section).

**Measures**

Students were asked to complete an assessment before the Introduction phase (general domain knowledge pre-test), before the Hypothesis generation phase (variable-related knowledge test), and after the Conclusion phase (general domain knowledge post-test). The general domain knowledge tests aimed to assess students’ level of knowledge about the domain before and after using the learning environment. The variable-related knowledge test was to check students’ knowledge about the specific variables to be learned about in the domain before they dived into the inquiry learning task.

**General domain knowledge pre-/post-tests**

A general domain knowledge pre-test assessed students’ prior knowledge of the domain and a general domain knowledge post-test assessed their domain knowledge acquisition. Both general domain knowledge tests were paper-and-pencil tests with 12 multiple-choice questions (four answer alternatives) each. The 12 questions addressed six aspects of the learning topic: sum of forces, friction, the balanced situation of a stationary object, the balanced situation of a moving object, the unbalanced situation of a moving object, conclusions based on Newton’s first law of motion. These six aspects were decided on together with the head physics teacher of the target school as covering the topic to be learned about from the learning environment. Two questions were included for each aspect. Both tests were checked by the head physics teacher to make sure all the questions were to the point, and understandable for the target students. The response to each question was scored either 1 or 0, and the maximum score on both tests was 12. Parallel pre- and post-tests were used, meaning that each question on the pre-test
had a corresponding similar question on the post-test. The parallel questions on the post-test were developed by paraphrasing the question, changing the pictures given in the question, or substituting different numbers into the question. The reliability results for the general domain knowledge pre-/post-tests were .55 and .62 (Cronbach’s alpha), respectively. This somewhat moderate reliability might be due to the general nature of the domain knowledge being assessed, such that items addressing different aspects of the learning topic were not measuring one and the same construct, but several sub-constructs under a general construct. Another explanation could be the small number of items involved in the tests. Due to the limited time allowed for the whole experiment, only a small number of items was used in each test.

**Variable-related knowledge test**

The variable-related knowledge test was intended to check the effect of the intervention as far as preparing students with knowledge of the variables in the domain, prior to their generation of hypotheses. It was a paper-and-pencil test with 6 open-ended questions. These 6 questions covered all of the variables introduced or to be explored in the learning environment. Three questions asked about the definitions of the three key variables: force, motion and friction. One question asked about the sub-variables that can be observed or measured for force (the magnitude or size of a force, and the direction in which the force is applied), and another asked about the sub-variables for motion (speed and direction of motion). The last question aimed to motivate students to think about all of the variables that could possibly influence the relation between force and motion. This test was also checked by the head physics teacher of the target school.

Students’ answers to the 6 questions were scored based on levels of correctness. The levels of correctness were inductively determined by reading a sample of the answers and developing a coding scheme (see Appendix B), which was then applied to the entire data set. For each of the questions, if a student offered no response or an irrelevant response to a question, 0 points were assigned. For the three definition questions, if a variable was explained by describing a phenomenon associated with the variable (e.g., motion is when something is moving), or by mentioning a partial characteristic of a variable (e.g., a force is a push), 1 point was assigned. If the meaning or essence of a variable was fully explained, then 2 points were awarded (e.g., motion is a change in the position of an object). With regard to the two questions about the sub-variables for force and motion, 1 point was given for each correct sub-variable mentioned in their answers (maximum 2 points...
Chapter 3

each). For the last question, 1 point was given for each variable listed in the coding scheme that was mentioned (maximum 6 points). The maximum score for this test was 16. A second rater coded 10% of the data and the interrater reliability of the coding scheme was $.94$ (intraclass correlation). The internal reliability of this test was $.61$ (Cronbach’s alpha).

Coding of inquiry processes

To detect whether different conditions had an effect on students’ inquiry processes, especially their hypothesis generation process, a coding scheme (see Appendix C) for inquiry processes was designed. This coding scheme covered four key inquiry processes: hypothesis generation, data recording, drawing conclusions, and final summary. Since students’ performance in all of these processes was highly related with the learning topic, the selection of the aspects to be coded for each process was based on both the characteristics of the learning process and the content of the learning topic. It should be noted that experimentation is also an important process in inquiry learning. However, since we did not have access to the data for students’ experimentation behavior in the PhET lab, the coding of students’ inquiry processes only covered the four processes mentioned above.

Hypothesis generation is central in our work and we judged students’ general performance on hypothesis generation and the quality of the generated hypotheses. The number of hypotheses generated and the number of different variables used in all of their hypotheses together (introduced as diversity below) were the two indicators for students’ general performance on hypothesis generation. To gain more detailed insight into the quality of the hypotheses, we scored the quality on two separate aspects, testability and informativeness (further introduced below). These two indicators were chosen based on the general framework in the assessment scheme used by van Joolingen and de Jong (1991), which coded students’ hypotheses on six aspects: well-formedness, description of hypothesized relation, relevance of a relation, complexity of a relation, precision, and generality. The first three aspects were covered by ‘testability’ and the latter three were integrated as ‘informativeness’ in the present study.

This coding scheme is similar to the coding scheme we used to code students’ inquiry processes in a prior study (Kuang et al., 2020). The main change lies in the coding of the informativeness of a hypothesis. Instead of generally coding the informativeness of a hypothesis from two aspects (variable selection and variable condition) as in the
previous coding scheme, the present coding scheme took a closer look at the informativeness of a hypothesis by specifying different levels for each of the three main aspects (variables, relations, and conditions) of a hypothesis. To explain the coding criteria for each level of informativeness for each aspect, an example hypothesis for each level was added to the present coding scheme (Appendix C).

**Coding of hypothesis generation**

*Number of hypotheses generated*

The total number of hypotheses generated by a student was counted as an indicator of students’ general performance on hypothesis generation.

*Diversity (number of different variables over all hypotheses)*

To explore the relation between force and motion, different variables related to force and motion need to be considered. The more valid variables considered in generating hypotheses, the more complete a picture a student can have of the learning topic. Hence, diversity in the variables used was coded by counting the number of different valid variables used across all of the testable hypotheses written by a student. Only the variables from testable hypotheses were counted and each variable was counted only once. For instance, in the hypothesis, “if applied force is greater than friction, then the speed of the object will increase”, there are three different valid variables: applied force, friction, and speed. The diversity number did not change if these three variables were repeatedly mentioned as well in other hypotheses.

*Testability*

Two aspects of the testability of a hypothesis were coded: first, the validity of the chosen variables and second, the testability of the presumed relation. All of the variables that could be observed or adjusted in the force and motion simulation lab were listed in the coding scheme. If the variables stated in a hypothesis could be found on the list, and the dependent variable was not the same as the independent variable, then this was regarded as valid ‘variable set’ and 1 point was given. A testable relation in a hypothesis meant that a condition of an independent variable was stated and a corresponding outcome condition of a dependent variable was also indicated. One point was assigned if a testable relation was identified in the hypothesis.

65
Informativeness

In this study, we determined informativeness of a hypothesis as a combination of coverage and specificity. Hypothesis informativeness was coded for each of the three main elements of a hypothesis, namely, the stated variables, the relation between the variables, and the stated condition. Based on an iterative viewing of a sample of students’ hypotheses, the possible levels of informativeness of these three elements were categorized. With regard to the informativeness of the variables, a rubric for levels from 0 to 4 was developed. A distinction between Level 1 and Level 2 was made based on the coverage of the variables. At Level 1 only one of the variables force and motion was present. At Levels 2 and 3 both force and motion were present, and the specificity of the hypothesis determined which level was assigned. For instance, a Level 2 hypothesis is “If the applied force is greater than the friction, then the object will move”; a Level 3 hypothesis is “If the applied force is greater than the friction, the speed of the object will increase”. In this case, the dependent variable in the Level 3 hypothesis was more specific than the dependent variable in the Level 2 hypothesis. The difference between Level 3 and Level 4 lay again in the coverage of variables. Here is an example of a Level 4 hypothesis: If sum of forces is not 0, the speed of the object will increase. Because the independent variable sum of forces covered more situations than only stating one possible situation with regard to applied force and friction (as in the example of a Level 3 hypothesis), this was coded as a higher level of informativeness. Concerning the informativeness of the relation between variables, a 0-2 rubric was used to assign the level. Involving an intermediate variable when stating the relation increased the informativeness of the stated relation (e.g., If the applied force increases, then the friction also increases (Level 1); If the applied force is greater than the friction, then the speed of the object will increase (Level 2)). As for the informativeness of the condition, a 0-1 level scale was used to indicate whether the initial state of motion was stated in the hypothesis or not. This was coded because the initial state of motion plays an important role in the relation between force and motion. If the initial state of motion was indicated in a testable hypothesis, Level 1 was assigned, otherwise Level 0 was assigned. The level equaled the number of points assigned. A more detailed description of the coding and more examples of each coding are presented in Appendix C.

Coding of data recording

Students’ performance on data recording was coded for three aspects: if the initial state of motion of the object was recorded, if there was a description of the condition of the independent variable, and if there was a description of the
outcome condition of the dependent variable. Students were given 1 point for each aspect they recorded for each hypothesis in the Observation tool.

**Coding of drawing conclusions**

Students’ conclusions on whether to accept or reject each of their hypotheses were coded by evaluating whether the conclusion was correct and whether the conclusion was data-based. If a student accepted a hypothesis that should be accepted or rejected a hypothesis that should be rejected, 1 point was assigned. If the student’s conclusion could be inferred from the recorded data or observation, another point was awarded.

**Coding of final summary**

The final summary concerned the student’s answers to the two open-ended questions concerning the two main conclusions from Newton’s first law of motion: force can change motion and motion can be maintained by force. Responses to these two open-ended questions were coded in a similar way. For instance, when coding the answer to the first question on the situations in which force(s) can change an object’s state of motion, if there was no response or the answer was totally irrelevant, 0 points were assigned. If a general situation in which force can change the state of motion was described (e.g., when force is applied, an object can move), 1 point was assigned. If one or more specific situations were mentioned in which force can change the state of motion (e.g., when the applied force is greater than the friction, an object can move), 2 points were assigned. If a conclusion based on specific situations was mentioned (e.g., when the sum of forces ≠ 0 or is > 0, an object’s state of motion will change), 3 points were assigned.

The first rater coded the inquiry processes of all students. A second rater was trained to use this coding scheme to code the inquiry processes of 12 students (10%). In terms of hypothesis generation, the interrater reliability coefficient reached .93 (intraclass correlation). The interrater reliability coefficient for the data recording process reached .77 (intraclass correlation), for drawing conclusions reached .83 (intraclass correlation), and for the final summary reached .93 (intraclass correlation).
Results

Results for knowledge tests

Table 3.2 summarizes the descriptive statistics for students’ performance on the three knowledge tests, by condition. Students were assumed to have low prior knowledge of the learning domain, since it had not yet been taught in their regular physics course. However, students on average earned slightly more than half of the maximum score in the general domain knowledge pre-test in all conditions, indicating that students already had basic understanding of the learning topic before the intervention.

Table 3.2
Descriptive statistics for scores on the three knowledge tests, by condition

<table>
<thead>
<tr>
<th>Maxi-</th>
<th></th>
<th>Condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>mum</td>
<td></td>
<td>D+E</td>
</tr>
<tr>
<td>score</td>
<td></td>
<td>(n = 29)</td>
</tr>
<tr>
<td></td>
<td>M (SD)</td>
<td>M (SD)</td>
</tr>
<tr>
<td>-------</td>
<td>--------------------</td>
<td>----------</td>
</tr>
<tr>
<td></td>
<td>General domain knowledge pre-test</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>General domain knowledge post-test</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>Variable-related knowledge test</td>
<td>16</td>
</tr>
</tbody>
</table>

The normality of the general domain knowledge pre-/post-test scores was confirmed by the results of the Shapiro-Wilk test ($p > .05$ for all the four conditions). But the Shapiro-Wilk test results for the variable-related knowledge test scores for conditions E and C were significant ($p = .02$ and $p = .03$, respectively), showing that the variable-related knowledge scores for these two conditions were not normally distributed. Hence, parametric tests were used for the analysis of the general domain knowledge test scores, and non-parametric tests were used for the variable-related knowledge test scores.

A one-way ANOVA was used to compare students’ general domain knowledge pre-test scores across conditions. The results showed that there were no significant
differences across conditions, indicating that students in the four conditions had equivalent prior knowledge to begin with; $F(3, 114) = 1.92, p = .13, \eta^2 = .05$. Paired-samples $t$-tests were used to examine knowledge acquisition from the general domain knowledge pre-test to the post-test by students in each condition. Students from the D condition were found to have a significant increase in their mean general domain knowledge scores from pre-test to post-test; $t(29) = 3.06, p = .01$, Cohen’s $d = .56$. No significant differences were found for the other three conditions; D+E condition: $t(28) = .32, p = .75$, Cohen’s $d = .06$; E condition: $t(31) = 1.61, p = .12$, Cohen’s $d = .29$; C condition: $t(26) = -.13, p = .90$, Cohen’s $d = -.02$. In other words, only the learning environment that provided students with domain information effectively facilitated students’ knowledge acquisition.

A mixed-design ANOVA was conducted to detect any effect of the different forms of support on students’ general domain knowledge gain from pre-test to post-test. The results yielded no significant interaction effect of experimental conditions and measurement time, showing that students’ acquisition of general knowledge of the domain was not significantly impacted by experimental condition; Pillai’s trace $= .06, F(3, 114) = 2.26, p = .09$, partial $\eta^2 = .06$. Given the small sample size, the data were re-examined by estimating a Bayesian repeated measures ANOVA to compare the fit of the data under the null hypothesis, compared to the alternative hypothesis. An estimated Bayes factor ($BF_{01} = 2.21$, null/alternative) indicated that the data were 2.21 times more likely to occur under the null hypothesis than under a model including an interaction effect of experimental condition and time the test was taken. This result provides anecdotal evidence in favor of the null hypothesis.

The differences across conditions in scores on the variable-related knowledge test were checked with a Kruskal-Wallis test. Given the multiple testing among the four conditions, a Bonferroni correction was applied to counteract the problem of multiple comparisons. The corrected significance threshold was 0.008. A significant difference was found ($H = 15.02, p = .002, \eta^2 = .11$), indicating that there was a significant effect of experimental condition on students’ knowledge of the variables. Pairwise comparisons of condition revealed that students from the D condition performed better than students from the C condition on the variable-related knowledge test and this difference was statistically significant ($p = .004$). Students from the D+E condition and E condition performed relatively the same as those in condition C ($p = .024$ and $p = 1$, respectively). No significant differences were found between the D+E, D, and E conditions.
Results for inquiry processes

Results for hypothesis generation

Table 3.3 shows the descriptive statistics for students’ general performance on hypothesis generation (number of hypotheses generated and diversity of the variables stated in the hypotheses). No significant effect of condition was found on the number of generated hypotheses; \( F(3, 114) = .45, p = .72, \eta^2 = .01 \); or on the diversity of the variables stated in the hypotheses; \( F(3, 114) = 1.44, p = .24, \eta^2 = .04 \).

The quality of hypotheses was compared on two aspects across conditions: testability and informativeness. All of the participants wrote more than one hypothesis. To control for differences in the total number of hypotheses generated, each student’s scores for hypothesis quality were averaged by dividing the total score for each coding item by the total number of generated hypotheses. In other words, the following comparison was based on students’ average performance on hypothesis generation, per hypothesis.

Table 3.4 presents the descriptive statistics for the average scores for each coding item. A Kruskal-Wallis test revealed no significant differences across conditions on the number of valid variables (\( H = 5.69, p = .13, \eta^2 = .02 \)), number of testable relations (\( H = 2.50, p = .48, \eta^2 = .004 \)), informativeness of variables (\( H = 4.16, p = .25, \eta^2 = .01 \)), informativeness of the relation (\( H = 2.42, p = .49, \eta^2 = .005 \)), and informativeness of the condition (\( H = 5.38, p = .15, \eta^2 = .02 \)).

Table 3.3
Descriptive statistics for the basic information on students’ hypothesis generation performance, by condition

<table>
<thead>
<tr>
<th></th>
<th>Condition</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>D+E ((n = 29))</td>
</tr>
<tr>
<td>Number of hypotheses generated</td>
<td>M (SD)</td>
</tr>
<tr>
<td>Number of different variables mentioned in the hypotheses (diversity)</td>
<td>4.97 (0.50)</td>
</tr>
<tr>
<td></td>
<td>4.45 (1.55)</td>
</tr>
</tbody>
</table>
Table 3.4
Descriptive statistics for the average scores for testability and informativeness of hypotheses

<table>
<thead>
<tr>
<th></th>
<th>Maximum score (per hypothesis)</th>
<th>Condition</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>D+E (n = 29)</td>
<td>D (n = 30)</td>
</tr>
<tr>
<td></td>
<td>M (SD)</td>
<td>M (SD)</td>
</tr>
<tr>
<td>Testability</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Valid variables</td>
<td>1</td>
<td>.89 (.25)</td>
</tr>
<tr>
<td>Testable relations</td>
<td>1</td>
<td>.83 (.28)</td>
</tr>
<tr>
<td>Informativeness</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Informativeness of variables</td>
<td>4</td>
<td>2.18 (.75)</td>
</tr>
<tr>
<td>Informativeness of relations</td>
<td>2</td>
<td>1.33 (.54)</td>
</tr>
<tr>
<td>Informativeness of condition</td>
<td>1</td>
<td>.29 (.26)</td>
</tr>
</tbody>
</table>

It was assumed that students’ variable-related knowledge could further impact their performance on hypothesis generation. Table 3.5 summarizes the correlations (Pearson correlations) between students’ knowledge of variables in the domain and their hypothesis generation performance. The results show that students’ knowledge of variables was significantly related to all measured aspects of their performance on hypothesis generation, except for their performance on informativeness of the condition stated in a hypothesis (that is, whether students identified the initial state of motion).

Results for subsequent inquiry processes

Considering that only testable hypotheses can be further tested by recording data and drawing conclusions, students’ performance on data recording and drawing conclusions was only coded when the hypothesis was testable. To control for the differences in the number of testable hypotheses, each student’s average scores for data recording and drawing conclusions per testable hypothesis were used for further analysis (see Table 3.6). The data for two students (one from the D+E condition and one from the C condition) were also removed from the following...
analysis, since none of their generated hypotheses was testable. Since the final summarizing questions were shown to students only once at the end of the learning environment, the total scores students earned from these questions were used for further comparison.

Table 3.5
Correlations between students’ variable-related knowledge and hypothesis generation performance (Pearson correlation)

<table>
<thead>
<tr>
<th>Hypothesis generation performance</th>
<th>Variable-related knowledge test score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diversity of variables</td>
<td>.28**</td>
</tr>
<tr>
<td>Valid variables</td>
<td>.23*</td>
</tr>
<tr>
<td>Testable relations</td>
<td>.27**</td>
</tr>
<tr>
<td>Informativeness of variables</td>
<td>.24**</td>
</tr>
<tr>
<td>Informativeness of relations</td>
<td>.27**</td>
</tr>
<tr>
<td>Informativeness of condition</td>
<td>.09</td>
</tr>
</tbody>
</table>

Note. ** = p < .01, * = p < .05. N = 118

Table 3.6
Descriptive statistics for the average scores for data recording and drawing conclusions and the total score for final summaries

<table>
<thead>
<tr>
<th>Main variables</th>
<th>Maximum score</th>
<th>Condition</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>D+E (n = 28)</td>
</tr>
<tr>
<td></td>
<td>M (SD)</td>
<td>M (SD)</td>
</tr>
<tr>
<td>Data recording</td>
<td>3</td>
<td>1.76 (.57)</td>
</tr>
<tr>
<td>Drawing conclusions</td>
<td>2</td>
<td>.79 (.53)</td>
</tr>
<tr>
<td>Final summaries</td>
<td>6</td>
<td>2.86 (2.34)</td>
</tr>
</tbody>
</table>

A Kruskal-Wallis test indicated no significant differences between conditions on data recording ($H = 5.22$, $p = .16$, $\eta^2 = .02$), drawing conclusions ($H = .17$, $p = .98$, $\eta^2 = .03$), and final summaries ($H = 5.45$, $p = .14$, $\eta^2 = .02$).
Discussion

The present study investigated the effects of presenting domain information together with or instead of offering exploratory practice for facilitating students’ knowledge acquisition, hypothesis generation, and other subsequent inquiry processes. The basic idea behind the study was the finding from experimental studies (Hattie & Donoghue, 2016; Lazonder & Harmsen, 2016) and from analyses of PISA 2015 data (L. Chen et al., 2017) that for fruitful inquiry learning, students should be prepared and have sufficient initial knowledge. This preparation can take place in the form of direct instruction or through letting students do preparatory exploration on their own. In this study, before entering into the inquiry process, participants were either provided with domain information together with an opportunity to explore a simulation-based representation of the learning domain (D+E condition), provided only with the same amount of domain information (D condition), provided only with the opportunity to explore the simulation-based representation of the learning domain (E condition), or provided with no additional support at all (C condition). Students with support were generally expected to perform better than students from the control condition. Providing students with domain information coupled with exploration was predicted to better enhance students’ performance on hypothesis generation, the quality of their hypotheses, data recording, and drawing conclusions, which would further lead to better knowledge acquisition than in the other conditions. The results instead support the idea that offering domain information alone helps students to know more about the variables in the domain at the start of the inquiry process (on the variable-related knowledge test) than students who received no additional support; there are no indications that exploration by itself helps students with this. We also saw that the scores on this variable-related knowledge test were (moderately) correlated with hypothesis quality measures. However, no effect of condition on the quality of generated hypotheses and the subsequent inquiry process was found and, possibly for that reason, there was also no effect of condition on improvement in scores on the general domain knowledge test. Providing domain information alone was the only condition in which students improved from pre- to posttest.

In the set-up of this study, we expected that providing students with domain information before the inquiry process would give them information about the variables per se. Providing domain information followed by exploratory practice
as a potential method to support students’ inquiry learning was based on positive findings in favor of presenting domain information before inquiry learning (Barzilai & Blau, 2014; Lazender et al., 2010; Wecker et al., 2007; Wecker et al., 2013), and it is also in line with principles of cognitive load theory (Martin & Evans, 2019). We also expected that giving them the opportunity to explore would (additionally) give them the opportunity to see the relations between variables and thus would give them a better view, necessary for inquiry, of the structure of the domain. We found indications that giving direct domain information had an effect on improving students’ knowledge about variables before generating hypotheses and fostering their knowledge acquisition from the learning environment, but we can conclude that there is no evidence that exploration before inquiry should be used as a way to support the inquiry process, despite indications from previous work that this could help students (DeCaro & Rittle-Johnson, 2012). For the time being, based on our data it cannot be advised to have students explore the simulation before moving to the actual inquiry process. However, we offered this exploration phase without too much guidance, and it may be the case that the exploration phase preceding the inquiry process also needs guidance, just as does the inquiry process itself. It was observed by the experimenter that several students kept increasing the amount of force and did not explore other options, or piled up person and boxes on the ground in the simulation lab during the exploration. It was reasonable to infer that some students may have played with the simulation lab to check out what they already knew rather than trying out the newly introduced variables or exploring what they did not know yet. Unsystematic exploration has been found to be ineffective in promoting learning benefits (Dalgarno et al., 2014).

Although it was not possible in the present study, analyses of students’ behavior patterns during the exploratory practice would be helpful to shed light on how we can support students’ exploration in the future. In a recent study, Newman and DeCaro (2019) examined different ways of exploring concepts related to statistical variance prior to direct instruction. They found that exploration before instruction was more effective than the reverse sequence, but also found that the more guidance was given in the exploration phase, the better the results. This could mean that in our case, more guidance while students were exploring the variables should have been included. We provided students with exploratory practice in an attempt to offer them an opportunity to explore variables in the learning domain to deepen their conceptual understanding. However, it might be the case that students may have played rather than explored during the exploratory practice. A
second consideration may be the sequence in which we offered domain information and exploration. We found that the D+E condition did not give better results than the D condition on its own. In light of the effectiveness of using the problem-solving prior to instruction (PS-I) method (Holmes et al., 2014; Loibl & Leuders, 2018; Loibl & Rummel, 2014a, 2014b) for fostering students’ conceptual learning, it would be interesting to investigate whether presenting domain information after exploratory practice could be an effective support measure for hypothesis generation in the future. The PS-I method is intended to activate students’ prior knowledge and intuitive ideas in the initial problem-solving or exploration phase, thus enabling students to realize their knowledge gap and motivating them to integrate new information obtained during the subsequent instruction phase (Loibl & Rummel, 2014a).

Another assumption that we worked from was that improved knowledge of variables would improve the quality of hypotheses and that this would propagate throughout the rest of the inquiry process. We found a moderate correlation between scores on the variable-related knowledge test and scores for the quality of hypotheses, which may indicate that knowledge of variables indeed coincides with quality of hypotheses. Despite the fact that there were no significant differences between conditions, the descriptive data also showed that the D condition generally had the highest scores on the different measures of hypothesis quality. However, this did not translate into better performance on the other inquiry processes of data recording, drawing conclusions, and final summaries. One missing link that may help to explain this could be the students’ experimentation behavior. In this experiment, we used a simulation in which we could not record the students’ actions, so we cannot draw any conclusions here, but it could well be that the students needed extra support in moving from a well-defined hypothesis to an informative experiment.

Our study also had a few restrictions and limitations. First, the internal reliability of the items in the knowledge tests was moderate, which increases the difficulty of detecting effects that may exist. Second, we may have had a testing effect. Students were tested three times in the present study: the general domain knowledge pre-/post-tests and the variable-related knowledge test. Tests can be sources of learning rather than assessment tools, since taking tests can improve retention of the tested content (Butler, 2010; Butler & Roediger, 2007; Roediger & Karpicke, 2006). It could be that students acquired knowledge about the domain from the tests they took, thus narrowing the difference between conditions in students’
knowledge acquisition. Third, students may have devoted different amounts of time to the different inquiry processes. Since the experiment was conducted during the regular time for physics courses, all of the students were allowed three 55-minute sessions to complete the inquiry learning processes and the assessment tasks. However, the fact that the experimental interventions were different across conditions may have caused differential time use for the inquiry processes. Students in the D+E condition presumably had the least time and those in the C condition had the most time for writing their hypotheses, recording data, and drawing conclusions. Although all of the students managed to complete all of the tasks within the three sessions, the difference in the available time for completing the inquiry learning processes might have limited the potential impact of the interventions. Fourth, the provision of terms for variables in the hypothesis scratchpad may have tempered the results. Students in all conditions were offered the same hypothesis scratchpad to write hypotheses, in which terms for variables and relations were provided. Although the hypothesis scratchpad does not offer any information on the meaning of the variables and the relations offered in the scratchpad had a general nature, there is still a chance that the given terms also benefited the students in the control condition. Fifth, students in the present study were assumed to know little about the domain, since they had not yet learned about this domain at school. But the general domain knowledge pre-test results showed that on average, students scored more than half of the maximum number of points on the test, for all conditions. Since students were randomly assigned to different conditions in this study, their levels of prior knowledge varied within each condition. This might also have narrowed the difference among conditions. Prior knowledge plays a paradoxical role in students’ learning (Dochy et al., 1999; Kalyuga, 2008). Students with little prior knowledge lack the knowledge framework on which to build new information, while students with moderate levels of knowledge who still have misconceptions would be resistant to changing their inaccurate knowledge. It would be interesting to investigate how the interventions would work out for students who actually have little prior knowledge. Finally, the power of this study may have been limited by the relatively small sample size.

In conclusion, this study pointed to some positive effects of providing students with domain information prior to going into an inquiry process. The study also made clear that, for the moment, there are no reasons to advise having students explore the simulation before moving to the actual inquiry process. In future studies, different ways to structure this exploratory process and a different
sequencing of exploration and direct instruction may be investigated to see if this would give different results. Providing more attention to the experimentation behavior of students as the link between hypothesis quality and their data interpretation and drawing conclusions from the inquiry is also an interesting avenue for future research.

References


Loibl, K., & Rummel, N. (2014b). Knowing what you don't know makes failure productive. *Learning and Instruction, 34*, 74-85. doi.org/10.1016/j.learninstruc.2014.08.004


Chapter 4
Providing adaptive domain information as support

Abstract

This study investigated the effects of providing domain information in an early stage of an inquiry process, together with an aligned hypothesis scratchpad, on inquiry learning, and hypothesis generation in particular. Participants were provided with basic domain information that was adapted to their prior knowledge (experimental condition) or received no introduction to the domain (control condition) before writing their hypotheses. Sixty-nine secondary school students from two countries were randomly assigned to the experimental or the control condition. These two conditions were compared on hypothesis generation, the subsequent inquiry processes of data recording and drawing conclusions, and knowledge acquisition. Results indicate that the supported students could specify more testable relations in their hypotheses, and could write hypotheses with higher levels of informativeness about variables, conditions, and relations. No differences between conditions were found on data recording, drawing conclusions, and knowledge acquisition. Limitations and directions for future research are presented.

* This chapter is based on:
Introduction

Students’ engagement in learning science is known to lead to deep learning (see e.g., Freeman et al., 2014). Inquiry learning emphasizes the engagement of students by involving them in self-directed investigations of a learning domain through a series of learning processes, including but not limited to posing questions, generating hypotheses, designing experiments, collecting data, and drawing conclusions. However, these are ambitious learning processes, which increase the cognitive demands on students. Leaving the inquiry process entirely up to students will not improve learning (DeCaro, DeCaro, & Rittle-Johnson, 2015; Mayer, 2004). A large body of research has suggested that students need supports to help them benefit from inquiry learning (Alfieri, Brooks, Aldrich, & Tenenbaum, 2011; D’Angelo et al., 2014; de Jong, 2019; Lazonder and Harmsen, 2016; Mayer, 2004).

Hypothesis generation is a crucial process in inquiry learning. Hypotheses act as logical steps in linking and guiding the inquiry processes that follow. A hypothesis refers to a tentative and testable proposition on a possible relation between variables in a given domain (de Jong and van Joolingen, 1998). In hypothesis generation, students organize their tentative answers to the questions posed, which can help students to become cognitively engaged with the question to be answered (H. J. Kim and Pedersen, 2011; Wenham, 1993). The subsequent inquiry processes are then purposefully organized and proceeded to test these tentative answers. Hypotheses can give guidance concerning what variables to actually consider and manipulate in experiments (de Jong, 2006; Swatton, 1992) and they can guide systematic data collection (M. Kim, Joung, & Yoon, 2009; Wenham, 1993). After that, when drawing conclusions about the given inquiry questions, hypotheses can also act as plausible directions for interpreting the findings (Guisasola, Ceberio, & Zubimendi, 2006). The pivotal role hypotheses play in inquiry learning explains why inquiry learning has also been described as a hypothesis-generating, hypothesis-testing, and hypothesis-revising procedure (Zimmerman, 2007).

However, hypothesis generation is also regarded as a complex process for students (Chang, Chen, Lin, & Sung, 2008; Njoo and de Jong, 1993). One problem is that noticing or identifying relevant variables to be considered in a hypothesis requires substantial conceptual domain-specific knowledge, which students may lack (de Jong and van Joolingen, 1998; H. J. Kim and Pedersen, 2011; Quintana et al., 2004). Another problem is that students find it hard to translate their intuitive ideas about
Providing adaptive domain information as support

a phenomenon into testable hypotheses (Quintana, et al., 2004). A third problem is that students fail to define possible relations between the variables (van Joolingen and de Jong, 1991). The major goal of this paper was to explore adequate support for facilitating students’ hypothesis generation.

Theoretical background

Many studies have shown that domain knowledge is one of the main determinants of the effectiveness of inquiry learning. Students’ domain knowledge and inquiry learning processes are intertwined with each other. Students need some surface-level or basic prior domain knowledge with which to start their inquiry process (Hattie and Donoghue, 2016; Morris, Croker, Masnick, & Zimmerman, 2012; Penner and Klahr, 1996). This basic domain knowledge includes knowledge of variables involved in the domain and their potential relations. It enables students to make sense of the given question and it may guide initial exploration by focusing attention on potentially important variables to investigate (Morris, et al., 2012; Penner and Klahr, 1996). A synthesis of meta-analyses by Hattie and Donoghue (2016) concluded that learning strategies for deep learning (including inquiry learning) must be accompanied by sufficient surface knowledge. Another meta-analysis by Schneider and Preckel (2017) also mentioned that prior domain knowledge is an important moderator factor influencing the effectiveness of active learning methods such as simulation-based learning.

The SDDS (scientific discovery as dual search) model developed by (Klahr & Dunbar, 1988) indicates the role of domain knowledge played in hypothesis generation. This model characterized inquiry learning as a search between two related problem spaces: hypothesis space and experiment space. The hypothesis space contains all the variables involved in a domain and all the possible relations among them. The experiment space contains all possible experiments that can be performed in the domain. According to this model, a student starts an inquiry learning activity with searching a hypothesis from the hypothesis space, and then continues searching experiments from the experiment space to collect and evaluate evidence to test the hypothesis. Prior domain knowledge and results from exploratory experiments are the two main knowledge sources of the hypothesis space, which suggest two possible ways to support students’ hypothesis generation. This study focused on offering students domain information as the support.
Several empirical studies have demonstrated that students’ prior knowledge of a domain can impact their hypothesis generation performance. Lazonder, Wilhelm, & Hagemans (2008) examined the type of investigative strategy students used in completing an inquiry learning task within either a familiar context (a concrete task) or an unfamiliar context (an abstract task). The concrete task provided participants with a scenario that they had some intuitive knowledge about. Participants were asked to investigate how each of four factors (training frequency, smoking, nutrition and a self-defined factor) impacted an athlete’s time in a 10,000-m race. The abstract task replaced the four factors with four geometrical shapes and invited participants to discover the impact of geometrical shapes on a numerical score. Results showed that more students with the concrete task started the investigation by generating hypotheses first, while students with the abstract task started by performing exploratory experiments first. The specificity of the hypotheses was found to be higher with the concrete task and remained relatively constant over time, while with the abstract task, the specificity of hypotheses increased gradually. Similarly, Beckmann and Goode (2014) compared students’ performance in solving a problem with four levels of familiarity. Participants were asked to learn about the causal structure of a linear system that contained three input variables and three output variables. The problem was described with different levels of semantic familiarity by using different cover stories and variable labels. They found that a semantically familiar problem context motivated students to generate a larger number of hypotheses.

However, in practical educational settings, students might have incomplete prior basic domain knowledge, which can impede rather than facilitate their learning (Alexander and Judy, 1988; Hmelo, Nagarajan, & Day, 2000). Domain knowledge plays an important role in guiding students’ attention to important features during problem solving (Morris, et al., 2012). Students with little or incomplete prior domain knowledge may be distracted by irrelevant features of the given problem or neglect relevant variables that are important to be considered in solving the problem. Inaccurate intuitive knowledge about variables may also hinder students from inferring possible relations between them (Dochy, Segers, & Buehl, 1999).

Some researchers have attempted to bridge students’ domain knowledge gap by providing students with domain information to facilitate their inquiry learning (Lazonder, Hagemans, & de Jong, 2010; Liew, Tan, & Seydali, 2014; Reid, Zhang, & Chen, 2003; Wecker et al., 2013; Zhang, Chen, Sun, & Reid, 2004). Reid, et al. (2003) and a follow-up study from Zhang, et al. (2004) provided a reference book
Providing adaptive domain information as support

button for students to use to access conceptual information about variables related to the learning domain during the learning process. They found that providing such interpretive support for students can foster more elaborate and deeper understanding of the explored domain. Lazonder, et al. (2010) offered help files in which the meaning of each variable within the domain, important specifics about the variable and the general direction of the effect, were presented. The results showed that students who had access to domain information before the inquiry task generated more specific hypotheses and acquired more knowledge compared to those who were not provided with domain information. Wecker, et al. (2013) organized a teacher-led whole-class introduction and discussion of the observable phenomena in the learning domain and a basic model that could be used to account for the phenomena. They found that such prior presentation of domain information had both a short-term and a longer-term positive effect on knowledge acquisition. In the study by Liew, et al. (2014), participants were provided with basic domain knowledge of an algorithm before working on a C-Programming algorithm task. They found that the frequency of reading domain information was significantly correlated with the total number of pairs of variables manipulated, which may imply that domain knowledge supports learners in involving more variables in their hypotheses.

Given that students differ in their levels of prior domain knowledge, a non-differentiated approach treating all students in the same way is not likely to work out well. Hsu, Gao, Liu, & Sweller (2015) investigated the effects of four different levels of instructional detail on facilitating learning for students with different levels of prior knowledge. Students worked on a learning task about positive and negative correlations in a simulation-based environment, and different levels of detail in the instructional guidance were provided by changing the level of concreteness of the guidance (e.g., ‘set the r value to 0.7’ or ‘set the r value to any value ranging from -1 to 1’) or by changing how the procedure was described (e.g., explain the procedure step by step or generally introduce the procedure with a bunch of text). The results indicated that higher levels of instructional detail benefited learning for students with low prior knowledge, whereas lower levels of detail facilitated learning for students with higher prior knowledge, which is in line with the expertise reversal effect (Kalyuga, Ayres, Chandler, & Sweller, 2003). Van Gog, Paas, & van Merriënboer (2008) examined the effects of studying sequences of worked examples that were product-oriented (only presented a problem solution) and/or process-oriented (additionally explained the rationale underlying the presented solution) on solving troubleshooting problems. They
compared the effect of providing product-oriented first or providing process-oriented first. The results revealed that studying process-oriented worked examples first fostered learning better than seeing product-oriented worked examples first. But after students became more knowledgeable over the course of the training process, continuing to study process-oriented worked examples hampered learning compared with continuing with product-oriented worked examples. A study by van Riesen, Gijlers, Anjewierden, & de Jong (2018) found that providing an experiment design tool for students with low prior knowledge fostered their knowledge acquisition, which was not the case for students with high prior knowledge. All of these findings suggest that the most effective approach requires adapting to students’ actual prior knowledge level. In a preceding study (Kuang, Eysink, & de Jong, under review), we compared the effect of providing students with domain information together with exploratory practice or providing either of them only for supporting hypothesis generation. We found indications that students who were provided with domain information before inquiry improved their knowledge of variables, and gained knowledge on the domain compared to students who received no domain information or students who explored the domain either as a self-standing activity or in combination with domain information. These results may be due in part to an approach that did not take students’ prior knowledge into account.

Thus, providing the right degree or amount of domain information tailored to students’ prior knowledge level may help to provide better supports for students. While it is difficult for a single teacher to cater for the needs of all students efficiently in classroom settings, computer-based adaptive tools provide opportunities to offset the limits to teachers’ available time and effort (Linn et al., 2014). The immediate and personalized supports provided by adaptive tools are also assumed to motivate students to deepen their understanding of the given information (Applebaum, Vitale, Gerard, & Linn, 2017). No studies have investigated the effect of adaptively providing domain information on facilitating students’ inquiry learning performance. In this study, the domain information is termed adaptive, meaning that a computer system is available to check students’ prior knowledge of a domain and adjust the levels of detail and amounts of domain information accordingly.
Research question and hypothesis

The research question of the present study was “Does providing domain information in an adaptive fashion facilitate students’ hypothesis generation, their subsequent inquiry processes, and their knowledge acquisition?” To answer this question, either adaptive domain information or no domain information was provided prior to the hypothesis generating and testing process. The effects on the subsequent inquiry processes and knowledge acquisition were also examined, because hypotheses play an important role in directing and guiding subsequent inquiry processes, and the ultimate goal of inquiry learning is to facilitate conceptual knowledge acquisition. It was hypothesized that students presented with adaptive domain information would perform better than those given no domain information as far as hypothesis generation and subsequent processes to test the hypotheses, and would acquire more conceptual knowledge about the domain.

Method

Participants

A total of 69 students (45.7% male), from one Italian secondary school (n = 50) and two Romanian secondary schools (n = 19), completed all three sessions of the experiment. The students were all from the second year of secondary school and between 11 and 14 years old (M = 12.19, SD = .60). Informed consent was obtained for all participants. Participants from each class in each school were randomly assigned to one of the two conditions: the adaptive domain information (AD) condition (n = 36, 52.8% male) or the control (C) condition (n = 33, 39.4% male).

Design

A quasi-experimental pre-test/post-test method was used to examine the effectiveness of providing adaptive domain information. Students’ conceptual knowledge on the target domain was assessed before and after the inquiry learning task with a pre-test and post-test respectively. Students’ performance on inquiry processes, especially on hypothesis generation was coded into scores for further comparison.
Learning environment

Students in both conditions worked on an online inquiry learning environment about a physics topic – Newton’s first law of motion. This learning topic was discussed and chosen together with the teachers of the target schools to fit in with their curriculum. The topic had not been taught to the students before the experiment took place. The inquiry learning environment was designed with the Go-Lab ecosystem, which is an integrated platform of rich open educational resources such as simulation labs, scaffolding apps, and authorizing and sharing tools (de Jong et al., 2021). A sequence of 10 phases was configured for students in the AD condition: (1) Pre-test, (2) About the study, (3) Introduction, (4) Preparation 1, (5) Orientation, (6) Hypothesis, (7) Preparation 2, (8) Investigation, (9) Conclusion, (10) Post-test. The Preparation 1 and Orientation phases were included to enable provision of the adaptive domain information; therefore, these two phases were not provided for students in the C condition, who encountered only the other eight phases. The learning phases for each condition were listed as tabs on the left side of the learning environment, which was shown to students in their local language to make sure all of the content was understandable. The next subsections further explain each of the 10 phases.

Pre-/Post-tests

These two phases were designed to assess students’ conceptual knowledge of the target domain before and after students engaged in the inquiry learning task. In these two phases, students were first introduced to requirements for completing the test (e.g., 10-minute time limit, answer the questions independently) and were then presented with the knowledge test. In each test, 12 items addressed the six key topics from the learning domain that the learning environment covered, namely, sum of forces, friction, the balanced situation of a stationary object, the balanced situation of a moving object, the unbalanced situation of a moving object, and the core of Newton’s first law of motion (when forces can change the state of motion and when the state of motion can be maintained). Each topic was measured by two multiple-choice questions (with four response alternatives). To make the pre-test and the post-test comparable in difficulty, the items in the post-test were created by paraphrasing the questions, changing the pictures, or substituting different values into the questions. Students’ answers for each item were marked as right (1 point) or wrong (0 points), meaning that the maximum possible score was 12. Students were given 10 minutes to complete the pre-test and the post-test each. This time limit was mainly based on the length and difficulty of the test. A
Providing adaptive domain information as support

physics teacher was consulted to ensure students could complete the test within the time limit. The reliability (Cronbach’s alpha) of the pre-test and post-test was .35 and .52, respectively, which a bit lower than the reliabilities obtained in the preceding study where these tests were used (.55 and .62 respectively; see Kuang et al., under review).

**About the study**

This phase aimed to offer an overview of the learning phases to students and to clarify the requirements for conducting the experiments (e.g., complete all of the tasks independently and carefully, ask for help if technical problems arise).

**Introduction**

In this phase, the learning domain of force and motion was briefly introduced by linking it to a daily situation, with both text and a picture provided. The learning goals of the inquiry task were also clarified at the end of this phase.

**Preparation 1**

This phase was included only for students in the AD condition, to check students’ prior knowledge of the domain, to which the domain information presented in the next phase was adapted. First, students were provided with the main question to be investigated (What is the relationship between forces and motion when we push an object on a straight track?). Then, students were asked to think about variables that they thought were important to be considered to answer the inquiry question. After that, a selection tool with both relevant (e.g., friction, speed) and irrelevant variables (e.g., gravity, density of the object) was provided (see Figure 4.1). Students were asked to select variables that they thought were important to take into consideration. To increase the chance of careful selection, students were reminded not to select a variable just because it sounded familiar, but to select a variable if they knew it plays a role in answering the inquiry question. A selected variable block would turn from light blue to dark blue (see Figure 4.1). Students were also asked to indicate general relations between variables by drawing lines between variables. This could be done by clicking the activated green button of a selected variable and dragging a line to another selected variable. Students’ basic knowledge did not need to be correct. The selected variables and indicated links between variables would be used as clues to provide appropriate domain information in the following phase. A three-minute instructional video on how this selection tool works was provided before presenting the tool.
**Orientation**

In this phase, students in the AD condition were provided with domain information. Based on students’ selections and indicated relations in the selection tool presented in the Preparation 1 phase, students were automatically provided with an introduction to the variables that was adapted to their prior knowledge. To be more specific, for each of the 10 variables shown in the selection tool, a student could be provided with a full introduction, a simple introduction, or no further introduction to the variable in this phase.

*Figure 4.1. An example of selected variables and general relations indicated between them in the selection tool*

The underlying logic for which level of introduction to be provided is laid out in brief in Table 4.1. Whether and how to offer domain information on each variable shown in the selection tool mainly depended on three aspects: whether a variable was selected or not, whether the variable was a relevant variable or not, and whether the student indicated relations between this variable and the other variables or not. Based on these aspects, four possible situations could arise. Situation 1: if a relevant variable was not selected, then we assumed that the student lacked basic knowledge about the variable and the system would provide a full introduction of the variable. Situation 2: if an irrelevant variable was not selected, then we assumed that the student aware of its irrelevance, and no further introduction to the variable was given. Situation 3: if a
relevant variable was selected, there were two sub-situations. First, if this variable was linked with one or more relevant variables, we assumed the student had basic knowledge about the variable. However, because this could be the result of guessing and also to remind students of basic knowledge about the variable, the system provided a simple introduction to the variable. Second, if this selected relevant variable was not linked with any other variables or was linked with any irrelevant variable, we assumed the student had incomplete or inaccurate understanding on the variable. In this case, a full introduction to the variable was provided. Situation 4: if an irrelevant variable was selected, then we assumed the student had inaccurate understanding of the variable, and a full introduction was presented.

The full introduction of a variable consisted of (a) introducing the variable by linking it to a daily-life situation with both text and a picture to help students think through or perceive the new content, (b) presenting the meaning of the variable with both text and an example picture, and (c) introducing a few specifics on the variable with text and/or pictures (e.g., the unit of force is called a newton, and it has the symbol N). If the variable was irrelevant, then instead of introducing specifics on the variable (c), the reason why the variable was not relevant was explained. The format of the simple introduction was: (a) present the meaning of the variable with text, (b) introduce a few specifics on the variable with text and/or pictures.

Table 4.1
The underlying logic on whether and how to offer domain information for each variable

<table>
<thead>
<tr>
<th>Selected or not?</th>
<th>Relevant variable or not?</th>
<th>Relations indicated or not?</th>
<th>Provide domain information or not?</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Not selected</td>
<td>Relevant</td>
<td>Not applicable</td>
<td>Full introduction</td>
</tr>
<tr>
<td>2 Not selected</td>
<td>Irrelevant</td>
<td>Not applicable</td>
<td>No further introduction</td>
</tr>
<tr>
<td>3 Selected</td>
<td>Relevant</td>
<td>The selected variable is linked with one or more relevant variables.</td>
<td>Simple introduction</td>
</tr>
<tr>
<td></td>
<td></td>
<td>The selected variable is not linked with other variables or is linked with any irrelevant variable.</td>
<td>Full introduction</td>
</tr>
<tr>
<td>4 Selected</td>
<td>Irrelevant</td>
<td>Any indicated relation or no relation</td>
<td>Full introduction</td>
</tr>
</tbody>
</table>

95
**Hypothesis**

In this phase, students were first introduced to the concept and format of a testable hypothesis. Then a worked-out example hypothesis was shown to illustrate the main idea and main elements of a testable hypothesis. After that, an instructional video was provided to show the students how to use the hypothesis tool to write their hypotheses. Then the main inquiry question on force and motion was stated, and students were asked to write at least five hypotheses related to the given question using the presented hypothesis tool. This hypothesis tool had two main sections: a term section and a hypothesis section. In the term section, terms for variables, relations, and conditions were listed in blocks, which students could use to build their hypotheses. A “type your own” term was also offered for students to type their own terms, if necessary. The hypothesis section was where students could drop the terms to complete their hypotheses. In line with the experimental set-up that students in the AD condition were introduced with relevant variables and clarified the irrelevance of several variables if needed while students in the C condition were not introduced with variables, students in the AD condition were provided with terms for all relevant variables, whereas students in the C condition were not provided with terms for variables.

**Preparation 2**

This phase was included to prepare students with information on how to use the PhET lab on force and motion (a simulation lab designed by PhET Interactive Simulations, University of Colorado Boulder, licensed under CC-BY 4.0. https://phet.colorado.edu.), with which they could test their hypotheses in the next phase. Five instructional pictures created by labelling clickable or adjustable options of the PhET lab were shown to students in this phase (see an example in Figure 4.2).

**Investigation**

In this phase, students were shown the hypotheses that they had generated in the Hypothesis phase and were asked to test their hypotheses by designing experiments in the Force and Motion lab and recording the data. Before beginning to work in the simulation lab, students were presented with several hints on experimental design (e.g., To test each of your hypotheses, set the condition of variables in the Force and Motion lab according to the hypothesis, and then observe what happens; reset the lab every time you start a new experiment) and on data recording (e.g., For each of your hypotheses, please record the settings of variables and your observed outcomes in the data table below, because you will
need the recorded data to draw conclusions later on; If you want to record change of motion, please record both the initial state of motion and the following state of motion to show the change). After seeing these hints, students were provided with a table to record data coming from their experiments for each of their hypotheses and the Force and Motion lab to use for their experiments.

![Figure 4.2. An example of an instructional picture showing how the Force and Motion lab works](image)

**Conclusion**

In this phase, students were first shown their hypotheses and recorded data. Then students were asked to draw conclusions on whether to accept or reject each of their hypotheses. In the end, students needed to summarize what they had learned about the relationship between force and motion by answering two open-ended questions: In what situation will forces change an object’s state of motion? And in what situation will an object maintain its original state of motion?

**Procedure**

Students in both conditions participated in three experimental sessions. All sessions took place during the students’ regular physics lessons and were led by the students’ physics teacher. Due to the COVID-19 situation, the experiment was conducted totally online. For standardization, the first author explained the experimental procedure to all the teachers involved prior to the first experimental session, and
gave them instructions in the form of a three-session lesson plan. The teachers were asked to follow the lesson plan, which contained details about the procedure for each session and highlights of what the experiment required. The length of a regular lesson differed across the schools, which meant that the length of the sessions differed. However, the total time did not differ much overall between schools, and since students were randomly assigned to the conditions within schools, total time did not differ between the two conditions. The procedure for each session and the time allowed for knowledge tests were still the same across all schools.

In the first session, after a brief introduction to the study, students were given a link to enter the learning environment. A specific link was given for each condition, without the students knowing this. After logging in, students started to work on the Pre-test phase. To avoid distraction from the subsequent learning phases, and to control the time for the test, only the Pre-test phase was visible for students when they logged into the learning environment for the first time. Students were allowed around 10 minutes to complete the pre-test; after that, the tab of the pre-test phase was hidden and the subsequent learning phases except for the post-test phase were available for them to work on. In the rest of the session, students worked on the subsequent phases.

The second and third sessions took place several days later, depending on the school’s lesson schedule, but within one week. The procedure for these two sessions was similar. Students were asked to take five minutes to review what they had done and continued from where they stopped in the previous session. The post-test was only shown to students when all the students in the class had completed the Conclusion phase. During post-testing, only the Post-test phase was visible for students. Students were also asked to complete the test within 10 minutes.

Coding scheme for inquiry processes

In order to compare students’ performance on inquiry processes, a coding scheme (see Appendix C) was developed to assess their hypothesis generation, data recording, drawing conclusions, and final summarizing processes. This coding scheme was used in an earlier study (Kuang et al., under review); in this study, there was interrater reliability of .77 to .93 (intraclass correlation) depending on the inquiry process assessed. It should be noted that experiment design is also an important process in inquiry learning. However, since we did not have access to the log files for the PhET lab, students’ experiments could not be coded for further comparison.
**Hypothesis generation**

Students’ performance on hypothesis generation was coded and compared on four aspects. Two were quantitative measures of the complete set of hypotheses generated (total number of generated hypotheses and diversity of variables used in hypotheses), while the other two were quantitative measures of each individual hypothesis (testability of the hypothesis and informativeness of the hypothesis).

The **testability** of a hypothesis was determined by two aspects. First, it was determined whether the variables were all on the target topic of force and motion and whether the dependent variable was not the same as the independent variable. If the student’s hypothesis met both of these conditions, it was assigned one point for identifying a “valid variable set”; otherwise, it received zero points for this aspect. Second, it was determined whether the condition of the independent variable(s) was mentioned and whether a possible outcome condition of the dependent variable was also mentioned. Again, if the student’s hypothesis met both of these conditions, it received one point, in this case for specifying a “testable relation”, and otherwise zero points were awarded for this aspect.

The **informativeness** of a hypothesis was coded separately for the informativeness of the three main elements (variable, relation, and condition) of a hypothesis.

**Informativeness of variables**

Five levels of informativeness of variables were distinguished based on the **coverage** and **specificity** of stated variables. An example of coverage is: if the independent variable(s) and the dependent variable(s) all concern force (e.g., If the applied force is equal to the friction, then the sum of forces is 0), then these variables are coded as less informative than variables that cover both force and motion (e.g., If the applied force is greater than the friction, then the object will move). With regard to specificity, general variables were regarded as less informative than specific variables, so that the hypothesis, If the applied force is greater than the friction, then the object will move is less informative than the hypothesis, If the applied force is greater than the friction, then the speed of the object will increase. In the case, the speed of the object in the second hypothesis is more specific than the general state of motion referred to in the first hypothesis. Informativeness of variables was scored from 0-4 points based on the assigned level, according to the scheme shown in Appendix C.
Informativeness of relation

Three levels of informativeness of the relation between variables were distinguished based on whether intermediate variables were involved in a hypothesis. An example of a direct relation is “If the applied force increases, then the friction increases”. An example of an indirect relation with an intermediate variable is “If the applied force is smaller than the friction, then the speed of the object will decrease”. A direct relation received one point, an indirect relation received two points. No response, or a non-testable hypothesis received zero points.

Informativeness of condition

The informativeness of the condition was coded based on whether the initial state of motion was stated in the hypothesis, as in the following example: If the applied force is larger than friction, then the speed of a stationary object will increase. When the initial condition was mentioned, one point was given; if not, the hypothesis received zero points for informativeness of condition.

Data recording

Data recording was coded on three aspects: if the initial state of motion of the object was recorded, if the condition of the independent variable(s) was recorded, and if the outcome condition of the dependent variable was recorded. For each aspect, students could receive one point, with a maximum of three points in total.

Drawing conclusions

Drawing conclusions was coded on two aspects: if the conclusion was correct (accept a hypothesis that should be accepted or reject a hypothesis that should be rejected), and if it was a data-based conclusion (if the conclusion can be inferred from the recorded data). One point was awarded for each aspect.

Final summary

Students’ answers to the two summary questions on the relationship between force and motion were coded according to four levels of correctness. The two questions were about when forces can change an object’s state of motion and when an object’s state of motion can be maintained. For example, in responses to the first question, zero points were awarded if there was no response, or if the response was totally irrelevant. One point was assigned if the answer generally mentioned that forces can move or stop objects (e.g., An object can move when forces are applied to it). Two points were awarded if the answer correctly mentioned one or
more specific situations when forces can change the state of motion of objects (e.g., When applied force is larger than friction, an object can move). Three points were given if the answer conclusively mentioned that when sum of forces $\neq 0$ or $> 0$, the state of motion will change. Students could receive a maximum of three points for each of the two questions.

**Results**

**Hypothesis generation**

The descriptive statistics for students' number of generated hypotheses and the diversity of variables in each condition are shown in Table 4.2. The results of a Shapiro-Wilk test for normality of distribution showed that the data for the number of hypotheses generated ($p < .01$ for both condition AD and condition C) and the diversity of variables mentioned in the hypotheses ($p = .01$ for condition AD and $p = .00$ for condition C) were not normally distributed. Hence, a Mann-Whitney U test was used to compare these two variables between conditions. The results showed that students of condition AD performed reliably better in the diversity of variables covered in their hypotheses ($U = 286.50$, $z = -3.77$, $p < .01$), but not better in the number of hypotheses generated ($U = 578.50$, $z = -.53$, $p = .60$) between conditions.

Table 4.2
*Number of hypotheses generated and diversity of variables, per condition*

<table>
<thead>
<tr>
<th>Main variables</th>
<th>Condition</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AD</td>
<td>C</td>
<td></td>
</tr>
<tr>
<td></td>
<td>($n = 36$)</td>
<td>($n = 33$)</td>
<td></td>
</tr>
<tr>
<td><strong>M (SD)</strong></td>
<td><strong>M (SD)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of hypotheses generated</td>
<td>4.97 (0.17)</td>
<td>4.97 (0.39)</td>
<td></td>
</tr>
<tr>
<td>Number of different variables mentioned in the hypotheses (diversity)</td>
<td>4.72 (1.54)</td>
<td>3.03 (1.83)</td>
<td></td>
</tr>
</tbody>
</table>

The number of hypotheses generated by students ranged from 3 to 6. To control for the difference in the total number of hypotheses generated, testability and
informativeness were compared based on students’ relative performance on hypothesis generation. This relative performance was calculated by dividing the sum score for each coded item over all hypotheses by the total number of generated hypotheses.

Table 4.3 shows the average score of each coded item for testability and informativeness, per condition. Because the data were not normally distributed, non-parametric tests were used for further comparison. It was found that students who were provided with adaptive domain information and terms for relevant variables performed better on average on indicating testable relations ($U = 410.00$, $z = -2.31$, $p = .02$), informativeness of variables ($U = 408.50$, $z = -2.24$, $p = .03$), informativeness of relation ($U = 330$, $z = -3.19$, $p = .001$), and informativeness of condition ($U = 307.50$, $z = -4.25$, $p < .001$), and these differences were statistically significant. The difference in average use of valid variables was not statistically significant ($U = 455.50$, $z = -1.82$, $p = .07$).

Table 4.3
Relative scores on testability and informativeness of hypotheses, per condition

<table>
<thead>
<tr>
<th>Main variables</th>
<th>Maximum score (per hypothesis)</th>
<th>Condition</th>
<th>AD ($n = 36$)</th>
<th>C ($n = 33$)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>$M (SD)$</td>
<td>$M (SD)$</td>
</tr>
<tr>
<td>Testability</td>
<td>Valid variable set</td>
<td>1</td>
<td>.85 (.23)</td>
<td>.70 (.35)</td>
</tr>
<tr>
<td></td>
<td>Testable relations</td>
<td>1</td>
<td>.79 (.28)</td>
<td>.60 (.37)</td>
</tr>
<tr>
<td>Informativeness</td>
<td>Informativeness of variables</td>
<td>4</td>
<td>1.72 (.69)</td>
<td>1.28 (.82)</td>
</tr>
<tr>
<td></td>
<td>Informativeness of relations</td>
<td>2</td>
<td>1.11 (.52)</td>
<td>.69 (.45)</td>
</tr>
<tr>
<td></td>
<td>Informativeness of condition</td>
<td>1</td>
<td>.18 (.23)</td>
<td>.01 (.05)</td>
</tr>
</tbody>
</table>

Subsequent inquiry processes

Considering that only testable hypotheses can be further tested by recording data and drawing conclusions, students’ performance on data recording and drawing conclusions was only coded when a hypothesis was testable. The data of one
student from the AD condition and six students from the C condition were removed from the following analysis because none of their generated hypotheses was testable. To control for differences in the number of testable hypotheses, each student’s scores for data recording and drawing conclusions were divided by the number of testable hypotheses for further analysis. It should be noted that the total scores instead of average scores students earned from the final two summary questions were used in comparing the conditions since these two questions were shown to students only once at the end of the learning environment. Data are presented in Table 4.4.

Based on the Mann-Whitney U test results, the difference in mean scores for final summaries (U = 301.50, z = -3.59, p < .001), was statistically significant and favoured condition AD, but was not significant for data recording (U = 457.00, z = -.22, p = .82) or drawing conclusions (U = 360.50, z = -1.60, p = .11).

Table 4.4
Relative scores for data recording and drawing conclusions and total score for final summaries, per condition

<table>
<thead>
<tr>
<th>Main variables</th>
<th>Maximum score</th>
<th>Condition</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>AD (n = 35)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C (n = 27)</td>
</tr>
<tr>
<td></td>
<td>M (SD)</td>
<td>M (SD)</td>
</tr>
<tr>
<td>Data recording</td>
<td>3</td>
<td>.97 (.84)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.91 (.71)</td>
</tr>
<tr>
<td>Drawing conclusions</td>
<td>2</td>
<td>.82 (.61)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.61 (.62)</td>
</tr>
<tr>
<td>Final summaries</td>
<td>6</td>
<td>2.61 (1.55)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.33 (1.53)</td>
</tr>
</tbody>
</table>

Domain knowledge pre-/post-tests

Table 4.5 shows the descriptive statistics for the pre-test and the post-test scores in each condition. The results of the Wilcoxon signed-rank test indicated that overall, students’ mean post-test score significantly increased from their pre-test score for all participating students (Z = -2.56, p = .01). This increase was statistically significant for the AD condition (Z = -2.05, p = .04), but not for the C condition (Z = -1.58, p = .12). This indicates that students in the AD condition learned from the inquiry learning task, but this conclusion could not be drawn for students from the C condition.
Table 4.5
Knowledge test scores, per condition

<table>
<thead>
<tr>
<th>Main variables</th>
<th>Condition</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AD (n = 36)</td>
<td>C (n = 33)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>M (SD)</td>
<td>M (SD)</td>
<td></td>
</tr>
<tr>
<td>Pre-test</td>
<td>5.81 (1.53)</td>
<td>5.58 (2.26)</td>
<td></td>
</tr>
<tr>
<td>Post-test</td>
<td>6.58 (1.95)</td>
<td>6.24 (2.12)</td>
<td></td>
</tr>
<tr>
<td>Gain score (post-test - pre-test)</td>
<td>.78 (2.22)</td>
<td>.67 (2.37)</td>
<td></td>
</tr>
</tbody>
</table>

The maximum score on the general domain knowledge pre-/post-tests is 12.

To examine whether condition influenced knowledge acquisition, a non-parametric ANCOVA with the rank of the post-test score as the dependent variable and the rank of the pre-test score as the covariate was performed. The ANCOVA results revealed that there was no effect of providing support on students’ knowledge acquisition; $F(1, 67) = 1.80, p = .19$.

**Discussion**

The main question investigated in this study was whether providing adaptive domain information for students can facilitate their inquiry learning process, especially hypothesis generation, and their knowledge acquisition. Providing domain information adapted to students’ prior knowledge was expected to prepare students with basic knowledge of relevant variables to be considered, which was anticipated to facilitate their hypothesis generation. Informative and testable hypotheses are an important guide for subsequent inquiry processes such as data recording and drawing conclusions. If students perform better on these subsequent inquiry processes, we may also expect this to lead to better knowledge acquisition. These conjectures were partially supported by the findings. Results indicated an advantage for providing adaptive domain (AD) information in facilitating students’ hypothesis generation and final summarisation process. Although the AD condition did not show a statistically significant advantage on data recording, drawing conclusions, and knowledge acquisition, the data revealed the expected trend for all of the inquiry processes and also for learning outcomes. This means
that, based on descriptive data that encompassed 11 measurements, the AD group always scored higher on processes and knowledge acquisition.

Results indicated that students in the AD condition considered a greater variety of variables in their hypotheses, indicated more testable relations in their hypotheses, and could write hypotheses with a higher level of informativeness with regard to variables, relations, and conditions. This finding confirms the important role that domain information plays in hypothesis generation (Beckmann and Goode, 2014; Lazonder, et al., 2008), and is consistent with previous studies (Lazonder, et al., 2010; Wecker, et al., 2013), which also found that presenting domain information was effective in fostering students’ performance on hypothesis generation. In classroom settings, providing students with the same amount and level of domain knowledge may be redundant for more knowledgeable students or insufficient for less knowledgeable students, which may hinder their learning (Kalyuga, 2007). Providing adaptive domain information, together with terms for variables, may have prepared all students in this condition with an adequate base of domain knowledge about the important variables to be considered, so that they could take more variables into account when writing their hypotheses and infer more testable relations between variables.

It should be noted that all effects were based on providing students with adaptive domain information together with terms for variables in their hypothesis tool. In the preparation phase, all relevant variables and several distracting irrelevant variables were introduced to students in the AD condition, whereas no variables were introduced in the C condition. To be in line with this and to avoid the possible influence of including variable terms for the C condition, all students in the AD condition were provided with terms for variables, whereas students in the C condition were not. However, hypotheses are never only about variables per se; the relations between variables are the key component of hypotheses. A study by Lazonder, et al. (2008) found that students who were provided with variables about which they had basic domain knowledge could write more specific hypotheses than students who had to use unfamiliar abstract variables. So, students need to know the meaning of the variables before they can infer possible relations between them. Because our students in the AD condition were also found to outperform their counterparts in the C condition in the level of informativeness of their hypotheses as far as relations and conditions, we assume that the adaptive domain information helped students to make sense of the given variable terms, which further facilitated their performance on hypothesis generation. To clarify
whether providing adaptive domain information without the variable terms could end up with the same effect, a control group with the adaptive domain knowledge but without the variable terms and a control group without the adaptive domain information but with the variable terms should be added in the experimental set-up for a future study.

Well-generated hypotheses were expected to benefit students with guidance in reasoning about subsequent inquiry processes, which is not fully confirmed by the results. No effects of providing domain information were found on data recording and drawing conclusions. A possible explanation is that students need more extensive support on data recording and drawing conclusions, which also present difficulties for students during inquiry learning (Mulder, Lazonder, & de Jong, 2010). Although students were provided with several hints on how to record data and a table in which to record their data, they might still be distracted by different variables simulated in the lab and find it challenging to decide what to record to test their hypotheses. Our students scored quite low on data recording in both conditions, which indicates students’ lack of data-recording skills. But to figure out what really happened and to make findings more interpretable, more in-depth research methods, such as self-reports, interviews, or observations could be added in further studies to understand students’ perceptions and behaviors. Furthermore, to avoid giving away information on the specific variables and relations to be considered in hypotheses, the provided hints and the data recording table did not specify particular variables for which the data should be recorded. However, it may be that support is needed to assist students in translating the conditions of variables or relations between variables stated in a hypothesis into recordable items. It would be of interest and value to figure out the right type of support to bridge between the stated hypotheses and appropriate data recording in future research. Drawing conclusions is based on the interpretation of the recorded data. The unsatisfactory performance on data recording may in turn have led to more challenges in drawing conclusions, which may have impacted the effect of the generated hypotheses on learning outcomes.

Despite the fact that students in the AD condition generated better hypotheses and performed better in answering summary questions about the domain, their performance in inquiry processes was not accompanied by significant higher knowledge acquisition. A possible explanation of this could be that the number of hypotheses generated and tested by students did not cover all aspects to be learned for the target learning topic. Due to the limited time allowing for completing the
inquiry tasks, including hypothesis generation, and the difficulty of writing hypotheses, it could be that students need to write and test more hypotheses in order to fully explore the learning topic.

One limitation of this study is that we did not compare the effect of providing adaptive domain knowledge and non-adaptive domain knowledge. Due to COVID-19 times, we only had access to a limited number of students and we decided to compare two conditions that were the most obviously apart. Future research could add a condition providing non-adaptive domain knowledge to students to further compare the difference. Another limitation is that, again due to the COVID-19 pandemic, the experiment was implemented fully online with students from four different schools in two different countries, in order to involve more participants. Although students in each class of each school were randomly assigned to the two conditions, the inherent differences between education in different countries might have some hidden impact on the results of the present study.

In conclusion, providing students with adaptive domain information and terms on variables was effective in facilitating their hypothesis generation. Students’ performance on hypothesis generation did not lead to significantly better learning outcomes, but descriptive data over a large set of measures showed a consistent trend in favor of the AD group in terms of processes and knowledge acquisition. Further research is needed to investigate how to steer students to translate hypotheses into guidance for subsequent processes, and in this way to potentially further improve students’ learning outcomes from inquiry learning.

References


Kuang, X., Eysink, T. H., & de Jong, T. (under review) Presenting domain information or self-exploration to foster hypothesis generation in simulation-based inquiry learning


Chapter 5

General discussion
Introduction

Inquiry learning emphasizes students’ active involvement in a rule-discovery activity through a series of learning processes, including but not limited to hypothesis generation, experimenting, data recording, and drawing conclusions. Being actively involved in an inquiry activity requires students to engage cognitively in the learning process (Chi & Wylie, 2014; de Jong, 2019). Hypothesis generation, one of the major processes of inquiry learning, is a good starting point for students to get cognitively engaged in the inquiry activity. The construction of hypotheses requires students to really think about the given problem and to relate what they know about the domain to the problem being addressed and organize their tentative ideas about a domain into testable relations among relevant variables (de Jong & van Joolingen, 1998; Kim & Pedersen, 2011). The relations between variables stated in hypotheses can further guide students’ attention to important points in subsequent inquiry processes and form the basis for concrete predictions of experimental outcomes. However, many students find it hard to write hypotheses on their own. Hence, the main goal of this dissertation was to explore ways to support students’ hypothesis generation in inquiry learning. To achieve this goal, three different types of support were designed, implemented, and investigated in the three studies in this dissertation.

A challenging task in supporting students is to balance between allowing the students freedom to explore and experience and restricting their actions by providing support for their explorations. On the one hand, in unguided inquiry learning, the uncertainty of the learning task and the freedom to decide upon and implement their own learning strategy can cause disorientation and learning obstacles for students, which makes it more likely that they will fail to benefit from inquiry learning (Lazonder & Harmsen, 2016; Mayer, 2004). On the other hand, providing support for students, while reducing uncertainty and guiding students’ attention to more productive parts of the learning process, might limit their freedom to explore during the learning process (Sun et al., 2022). Because there is a danger of over-restricting students when providing support, providing them with just enough support is the ultimate goal in assisting students. The three types of support investigated in this dissertation were three attempts to approach this goal.
The design of the studies and main findings

Study 1 started with providing students with a form of partial directive support — partial hypotheses — as the support. Directive support guides students in a certain direction (de Jong & Lazonder, 2014; de Jong & Njoo, 1992). Earlier research showed that providing a list of predefined hypotheses, a type of directive support, benefitted students’ inquiry learning process but not their knowledge acquisition (Njoo & de Jong, 1993b). A limitation of this directive support was that the fully predefined hypotheses limited students’ freedom to express their own ideas. Inspired by the completion strategy used by van Merriënboer (1990), who provided partial solutions for students in order to limit the task students need to do and to provide a referent example for students to use when proceeding with the learning task, Study 1 of this dissertation investigated the effect of providing partial hypotheses on students’ hypothesis generation. A partial hypothesis included a stated condition of an independent variable in the first half of the hypothesis, leaving students to complete the hypothesis by indicating a related dependent variable and a relationship between them. All students worked with a hypothesis generation tool, or hypothesis scratchpad, in which a set of terms to build hypotheses with was provided for the students. Students in the control condition had to build complete hypotheses themselves. Students in the experimental condition were provided with partial hypotheses that they had to complete. Results indicated that students who were provided with the partial hypotheses generated more informative hypotheses, performed better at data recording, and acquired more knowledge than students in the control condition.

Although providing partial hypotheses was found effective in improving students’ hypothesis generation and knowledge acquisition, the provided partial hypotheses still limited students’ freedom to explore what they wanted to investigate. Hence, Study 2 moved one step further by providing non-directive support, that is, support that could help students to generate their own hypotheses. The two knowledge sources for hypothesis generation suggested in the SDDS (scientific discovery as dual search) model (Klahr & Dunbar, 1988) shed light on possible directions of support for hypothesis generation. According to this model, hypothesis generation is either based on students’ prior knowledge, or based on exploratory experiences prior to hypothesis generation. Therefore, in this study, we investigated whether providing domain information together with exploratory experience, or either of them separately could facilitate students’ hypothesis
generation comparing to not having support. Results indicated that providing students with domain information alone helped to foster their knowledge of variables before generating hypotheses, and was the only condition in which students improved from pre- to posttest. There were no differences between conditions on the quality of hypothesis generation, subsequent inquiry processes, or knowledge acquisition.

Support that is effective in general may not benefit every student equally; similarly, support that is ineffective in general may not turn out to be useless for every student. To provide more appropriate support for students, Study 3 investigated whether making the domain information provided adaptive to students’ prior knowledge level would be a way to facilitate their hypothesis generation. Students were either provided with domain information that was adapted to their prior knowledge or received no introduction to the domain before writing their hypotheses. Results indicated that the supported students could specify more testable relations in their hypotheses, and could write hypotheses with higher levels of informativeness on variables, conditions, and relations compared to the students who did not receive support. No differences between conditions were found on knowledge acquisition. In this study, due to the COVID-19 pandemic, we only had access to a limited number of students and we could not include a condition in which students received non-adaptive domain information.

The general conclusion of the three studies are as follows:

- Providing partial hypotheses for students is an effective way to foster students’ performance in hypothesis generation, and also knowledge acquisition;
- Providing domain information alone helped students to know more about relevant variables, which did not further result in better performance on hypothesis generation;
- Providing exploratory practice did not show added value in facilitating hypothesis generation and knowledge acquisition;
- Providing adaptive domain information was found to benefit the quality of hypothesis generation, but not knowledge acquisition.
Discussion

This section presents a few reflections on the results, highlights and discusses several overarching methodological issues, and summarizes a few opportunities and challenges in doing research in times of a pandemic.

Reflections on the results

A frequent finding in the literature is that support facilitates learning processes, but not knowledge acquisition (e.g., Saab et al., 2007; van Joolingen & de Jong, 1991b). This work is no exception. In Study 3, the results indicated that providing adaptive domain information can foster better performance on hypothesis generation, which did not further lead to better knowledge acquisition. Van Joolingen et al. (2007) summarized the possible effects of support by introducing two levels: the first-order effect — better performance of learning processes and the second-order effect — better knowledge acquisition. They offered two reasons as to why the first-order effects might not lead to a second-order effect. The first potential reason is that improved learning processes need more time in order to contribute to knowledge building. The second reason could be that students focus more on completing the supported learning processes without using these processes for deeper learning. These two reasons provide insights for explaining the results of Study 3. In our case, students’ knowledge of the learning domain was measured right before and after the inquiry learning process with a pre-test and a post-test. A possible explanation as to why better performance on inquiry-learning processes including hypothesis generation did not result in better knowledge acquisition could be that students need more time to digest what they have experienced in the inquiry-learning processes and then transform their results into new knowledge. Another explanation could be that the supportive measure presented to students diverted their attention away from learning from the inquiry process and towards completing the required inquiry process, such as writing their hypotheses. This means that they might have spent most of their efforts on understanding and completing the tasks for each inquiry process, which could have hindered them from making use of the inquiry processes to learn more about the domain.

Another striking finding of this dissertation is that, contrary to our expectation, the non-directive support in Study 2 and Study 3 turned out to be relatively ineffective in fostering hypothesis generation compared with the partial directive support in
Study 1. The three studies in this dissertation aimed to explore ways to support students’ hypothesis generation, and also to explore what could be just enough support for students. The findings of the three studies seem to imply that providing partial directive support (partial hypotheses) was the appropriate level of support that allowed the target group of students to express their ideas when completing the incomplete parts, and did not constrain them too much at the same time. The three types of support were examined in three separate studies, but the age group of the participants, the learning domain, their prior knowledge of the domain, the general set-up of the inquiry learning task, the sessions allowed to complete the inquiry task, and the data analysis method were generally the same. Although further studies are needed to compare the effectiveness of the three types of support, it is reasonable to infer that non-directive support may be too demanding for students who have relatively low prior knowledge of the domain and limited or no experience in hypothesis generation and inquiry learning. Students at this level need more guidance to direct their attention to informative hypotheses.

However, it should also be noted that although the non-directive support in Study 2 and Study 3 did not achieve the desired effect, this does not necessarily mean that partial directive support is always better than non-directive support. Future research is needed to investigate if providing domain information, exploratory practice, or adaptive domain information can be beneficial for students who are more skilled in hypothesis generation, data recording, and drawing conclusions. It is also important to realize that one type of support may not benefit all students equally. Future research is required to investigate how to dynamically assess students’ knowledge level and skill level in order to automatically provide needed support for hypothesis generation on a just-in-time basis.

**Methodological issues**

**Issues concerning the tool design**

In this dissertation, we wanted to measure the effect of a specific intervention on inquiry processes and knowledge acquisition. This also implied that we had to design the tools in the environment according to the intervention, which concerned more specifically the design of the hypothesis scratchpad. Comparing the experimental set up of the three studies, apart from the type of the support, a major difference is that in Study 1 students were required to complete and test one
hypothesis at a time, whereas in Studies 2 and 3, students were asked to write and test all their hypotheses together. This difference was determined by the type of the support provided to the students. As mentioned earlier, in Study 1, students were provided with partial directive support—partial hypotheses. The partial hypotheses were provided to students one by one in order to gradually guide students in the intended direction. And the effect of the support was also assessed per hypothesis. In Studies 2 and 3, the non-directive support, domain information and explorations aimed to prepare students to generate their own hypotheses. Here we asked students to generate all of their hypotheses at one time.

Several considerations led us to ask students to generate all of their hypotheses together in Studies 2 and 3. First, we assumed that students, after having seen domain information and/or exploring the lab, had more than one hypothesis in mind. Therefore, it seemed natural to ask them for all of their hypotheses at once; we also thought they might forget some of their hypotheses if they started testing them one by one. Second, if students were asked to write one hypothesis at a time and also to investigate them one by one, this would mean that after testing their first hypothesis, they would use the simulation lab for investigation and gain knowledge before going on to test their following hypotheses. This would interfere with the experimental intervention involving providing domain information and exploratory practice.

This difference between Study 1 and Studies 2 and 3 in the number of hypotheses students were asked to generate at a time could perhaps also partly explain the result that the support in Study 1 was effective in improving knowledge acquisition, but the support in Studies 2 and 3 was not. Inquiry learning was new for students who participated in the studies in this dissertation, and they might have needed some experience with the hypothesis generation process as well as other subsequent inquiry processes in order to use them more effectively. If they had to generate all hypotheses at once, they had less chances to practice the whole process compared to shorter hypothesis generation and investigation cycles, and they might generate less informative hypotheses.

In our design, it was a dilemma that from the standpoint of a student’s complete inquiry learning experience, it would make more sense to test one hypothesis at a time, but from the standpoint of our experiment, minimizing the possible impact caused by extra experience in inquiry learning process was necessary. In future research, if students are given more time for engaging in inquiry within a domain,
it would be interesting to investigate the effect of gradually fading support for hypothesis generation in cycles of inquiry learning activities.

**Issues in the coding of hypotheses generated by students**

The main focus of this dissertation was to support students’ hypothesis generation and therefore, the assessment of students’ hypotheses was one of the major analyses done. Not only is hypothesis generation a difficult task for students, the assessment of hypotheses is also a challenging task for researchers. In developing an assessment scheme for hypotheses, we had to take into account characteristics of the hypotheses, while also considering these characteristics within the context of the specific domain involved. Our assessment scheme evolved throughout our experiments; in this section, we reflect on our considerations during its development.

The coding scheme for hypotheses that was developed in this dissertation is based on previous work. Quinn and George (1975) developed a 0-5-point hypothesis quality scale with five levels: 1) it makes sense, 2) it is empirically based, 3) it is adequate (scientific explanation relating at least two variables), 4) it is precise (a qualification and/or quantification of the variables), 5) it states a test. This coding scale was the first to cover different aspects of a hypothesis, but the differences between the levels were not straightforward enough, in our view. Van Joolingen and de Jong (1991b) assessed students’ hypothesis generation with more detailed and distinct aspects, including the number of well-formed hypotheses, the number of different variables and relations used in the hypotheses, and the number of relations classified according to level of preciseness (very global, qualitative descriptive, and conditional relations). Lazonder et al. (2009) zoomed in on one major characteristic of a hypothesis—specificity; they classified students’ hypotheses as fully-specified, partially-specified, and unspecified hypotheses, according to the level of domain specificity. Kim and Pedersen (2011) coded three aspects of students’ hypotheses: whether it relates to the given problem, whether it is a logical statement, and whether it is generated based on multiple pieces of evidence. Finally, in order to provide automated feedback on students’ hypotheses, Kroeze et al. (2019) developed a set of criteria that were based on the presence of certain elements or words in a hypothesis, for instance, contains at least two variables, contains a modifier, manipulates exactly one independent variable, and so forth.
Inspired by the above-mentioned work, we identified two main aspects of a hypothesis to be measured: the first is whether the hypothesis is testable, the second regards the quality of the content (variables, relations) in terms of the domain involved. The main question here is, does a hypothesis potentially contribute to the advancement of knowledge in the domain involved?

For testability, we developed the following definition: A hypothesis is testable when a relation between variables is stated in the hypothesis. To specify a relation between variables, a hypothesis needs to include the following five elements: 1) one or more independent variables (e.g., the applied force and friction), 2) the (relative) value of the independent variables (e.g., if the applied force is larger than the friction), 3) one or more dependent variables (e.g., speed), 4) the expected outcome value of the dependent variable (the speed will increase), and 5) all four of the above-mentioned elements should be from the given domain to be explored (so no variables or relations that are outside the given domain). Including each of these five points separately as coding items could be a valid coding scheme for the testability of a hypothesis. However, along with being valid, a coding scheme should also be efficient in terms of application (Garrison et al., 2006). Hence, in the coding scheme used in this dissertation, the five different coding elements for testability were condensed into two elements: valid variables and a valid relation between them. The validity of variables was further detailed by stating that all variables should be from the domain and the dependent variable should be different from the independent variable(s). The relation was regarded as valid if a value of the independent variable was stated and a value of the dependent variable was also indicated. The decision to condense the testability elements into these two elements was based on the fact that the coding task needed to be done manually. If new technology such as automated coding can be used and efficiency is not a problem, then all five of the specific elements can possibly be used for coding.

The second aspect of a hypothesis is the quality of its content. In Study 1, we used two coding elements (variable selection and variable focus) for quality. One point was assigned for variable selection if the hypothesis stated the relationship between force and motion rather than only either one of them. Another point was assigned for variable focus if the hypothesis focused on a balanced force situation, which was the target domain focus to be learned. At that point, we labelled this coded aspect the complexity of the hypothesis, which had a maximum score of two. This score was rather unrefined but straightforward.

121
In Studies 2 and 3, we took a closer look at the quality of students’ hypotheses. First, we changed the term complexity to informativeness, because we later realized that complexity can be interpreted as complication or difficulty, and this may have resulted in the misunderstanding that a more complicated hypothesis would be a better one. More importantly, we added the levels of informativeness of variable (four levels), relation (two levels), and condition (1 level) to the coding scheme. The classification of levels of informativeness of variables has a specific challenging issue which is to achieve a balance between the degree of specificity and the depth. There are situations when the more specific a hypothesis is, the better. For instance, to investigate a mathematical relation between two variables A and B, the relation stated in a hypothesis can be very general (A and B are related), or less general (If A changes, then B changes), or it may specify a direction of the relation (If A decreases, then B decreases), or specify the relation model (there is a linear relation between A and B), or specify the quantitative equation between the variables (A = 3*B; van Joolingen & de Jong, 1991a). In that case, the informativeness of a relation can always be classified in terms of its specificity. But in other situations, after being specific initially, at some point later on, the hypotheses should become general or abstract to form a generalizable rule or theory about the domain (Lawson, 2005). The learning domain of force and motion used in the studies in this dissertation fits the latter situation. In our case, students were asked to investigate the relationship between force and motion. When students were gradually specifying the type of force (applied force and friction) and the characteristics of motion (speed and moving direction), the hypotheses could become increasingly specific. But the learning goal was that students could finally summarize a relation between a general variable about force (sum of forces) and a general variable about motion (motion state of an object). A hypothesis mentioning the sum of forces would not be that specific, but it would be more effective for understanding the domain. But it should also not be the case that all hypotheses involving the general variable (sum of forces) should be scored as having higher level of informativeness. For instance, if the applied force is equal to the friction, then the sum of force is 0. Since all of the variables included in this hypothesis were about force, and the target question was about both force and motion, this type of hypothesis was scored as having a lower level of informativeness (coded as Level 1).

The coding scheme for hypotheses was revised several times; the main change between the one used in Study 1 and the one in Studies 2 and 3 was that the levels of informativeness of variables, relations, and conditions were included in Studies
2 and 3, but not in Study 1. We think that, over the experiments we conducted, we developed a comprehensive and informative coding scheme that is transferrable to other student levels and domains, but which, of course, needs to prove its usefulness in practice.

**Issues in the coding of subsequent inquiry processes**

The notion of inquiry learning shifts the emphasis of science education from instructing science content to engaging students in the process of constructing knowledge. The main processes involved in an inquiry-learning activity have been thoroughly discussed (Njoo & de Jong, 1993a; Pedaste et al., 2015; Quintana et al., 2004), while the assessment of the inquiry process remains an open question. Although knowledge acquisition is, with no exception, the ultimate learning goal of inquiry learning, assessing inquiry processes can be informative for knowing what students did and detecting what students needed. Some studies have included the assessment of one or more inquiry processes as part of their outcome measure (Lazonder et al., 2010; Roll et al., 2018; Zacharia & Anderson, 2003), yet the assessment methods differed among studies and there is a lack of commonly used assessing models or tools. The establishment of a transferable assessment model is quite an ambitious task, which is beyond the scope of this dissertation. The coding scheme for inquiry processes developed in this dissertation suggests a way to take the coherence of different inquiry processes into account when assessing the processes, which may contribute to the effort of reaching an assessment model.

The assessment of the subsequent inquiry process was hypothesis-oriented in this dissertation. This was not only because hypothesis generation was the focus of our research, but also because the hypothesis serves as good glue between the subsequent inquiry processes. Because the relations between variables stated in hypotheses guide the inquiry processes that follow, an inquiry learning activity has also been described as a hypothesis-generating and hypothesis-testing procedure (Zimmerman, 2007). Therefore, making the hypothesis central was a deliberate choice directed at making the assessment of the inquiry process coherent. In particular, the coded aspects of data recording and drawing conclusions were set to be hypothesis-oriented in this dissertation. For instance, if the condition of the independent variable was described in data recording, one point was given. If the outcome condition of the dependent variable was described, another point was given. The assessment of subsequent inquiry processes would have been more
complete if we could have had access to the data from students’ experiment activities, but we did not have access to log-files of their experimentation behavior.

Opportunities and challenges of doing research in times of a pandemic

The Covid-19 pandemic has impacted the way we do almost everything, including research. The data collection in Study 3 took place during the global pandemic, and encountered both challenges and opportunities as a result. The lockdowns and restrictions from the pandemic made online learning during the pandemic the mainstream form of education, which made it easier to “sell” our online inquiry learning activity to teachers, parents, and students. That is why a cross-countries study came into view. As an additional benefit, involving students from different schools and different countries could help us to double-check the outcomes of the same experiment, which could help to increase the reliability and generalizability of the results. In addition, as accessing online education became normal for both teachers and students, they were prepared to participate in online learning activities in terms of basic computer-related skills, online-related facilities, and regular sessions for online learning.

Of course, challenges also emerged along the way. Although students regularly followed online lessons due to the pandemic-related restrictions, the researcher-led experiment was changed into a teacher-led experiment because of the privacy restrictions on the access to join the online lesson. This change meant that we had much less control over the experiments. It also increased the workload in communicating all of the detailed requirements and procedures before the experiment and students’ performance and progress between the experimental sessions. Furthermore, the communication with the teachers had to be done with video calls and emails, which was less effective compared with face-to-face communication. Another challenge was that although we could reach more students to be participants, the drop-out rate was quite high. There were two major reasons for this. The first one that some students tested positive for Covid-19 or were found to be contacts of Covid-19-infected persons during the study. The second explanation was that the difference of location between the teacher and the students made it difficult for teachers to monitor and provide in-time explanations of the requirements or procedures if needed. The increased flexibility in studying from home may have resulted in less engagement in the online learning task, especially when students studied from home.
General conclusion

Hypothesis generation enables us to make sense of a problem by organizing what we know about a domain into relations between variables, which can further guide what to investigate in experiment design, what to record in data recording, and what to interpret in drawing conclusions. However, generating hypotheses is not an easy task for students. This dissertation aimed to explore effective, and if possible, just-enough support to facilitate students’ hypothesis generation. The three studies in this dissertation provided insights into how students can be supported in hypothesis generation and how students’ hypotheses can be assessed. The results suggest that providing partial directive support (partial hypotheses) is an effective support for students. Providing adaptive domain information led to generation of more testable and informative hypotheses, as expected, but failed to foster students’ knowledge acquisition. To optimize the possible effect of non-directive support, future research should examine the effectiveness of providing (adaptive) domain information and/or exploratory practice before hypothesis generation for students with different levels of knowledge and inquiry skills. In order to provide students with adequately individualized support, it is necessary first to investigate what support works better for students having a specific level of knowledge and inquiry skills.

References


English summary
Introduction

Inquiry learning in science enables students to construct conceptual understanding of a learning topic by actively involving them in a process of exploration and discovery. Inquiry learning has the potential for supporting students’ acquisition of both science content and science processes, but this cannot be achieved merely by having students explore in an inquiry learning context on their own. Students can encounter difficulties in all phases of inquiry learning, including hypothesis generation, experiment design, data recording and interpretation, and drawing conclusions. Support is needed to guide students through the complex processes involved in inquiry learning before they can benefit from it.

In this dissertation, we are concerned with the support of one process involved in inquiry learning — hypothesis generation. Students’ generation of hypotheses avoids superficial task processing in an inquiry learning activity, and elicits their cognitive engagement in the inquiry learning process. The construction of hypotheses requires students to think mindfully about the problem being addressed, relate the given problem to what they know, and organize their tentative ideas into relations between variables. The generated hypotheses can further direct what to consider in experiment design, what to record in data collection, and can shape students’ conclusions. However, students frequently experience difficulties in hypothesis generation. According to the literature, some students do not know what a testable hypothesis looks like and find it hard to generate initial hypotheses, some have difficulties in identifying relevant variables, and some fail to specify relations among variables.

This dissertation thus aimed to investigate possible ways to support hypothesis generation by students during inquiry learning. Finding the right level of support is a tricky balancing act. Offering too much support limits students’ freedom to make proactive decisions, while offering too little support leaves students to struggle with demanding tasks. Hence, the three experimental studies involved in this dissertation attempted to explore the provision of just-enough support for students’ hypothesis generation.
Overview of the studies

All studies in this dissertation were conducted with secondary school students (11-14 years old) and involved three experimental sessions during the students’ regular physics lessons. The online inquiry learning environments used in all studies were designed with the Go-lab ecosystem (https://www.golabz.eu), and concerned the same physics domain – force and motion. This learning topic was discussed and chosen together with the physics teachers of the target schools to fit in with their curriculum. In all cases, we made sure that the topic had not been taught to the target students prior to the experiment. The basic structure and the problem to be solved that was presented in the learning environments were roughly the same for all studies, but the exact tasks making up the inquiry processes differed per study depending on the research questions.

Students’ hypothesis generation was done by means of a hypothesis tool called the hypothesis scratchpad. In the standard version of the hypothesis scratchpad, the three main elements of a testable hypothesis (variables, relations, and conditions) were provided to the students as terms, which students could drag and drop to build their own hypothesis. A reusable “type your own” term was also provided for students to bring in their own ideas. The exact configuration of the hypothesis scratchpad differed slightly between studies based on the type of support provided in each study.

The effect of the provided support in all studies was examined in terms of students’ knowledge acquisition and students’ performance on inquiry learning processes, especially on hypothesis generation. Considering that the relations between variables stated in hypotheses are an important reference for subsequent inquiry processes, including data recording and drawing conclusions, we investigated the effect of the provided support not only on the generation of hypotheses, but also on the subsequent inquiry processes. Coding schemes were developed to assess students’ inquiry learning processes. Knowledge acquisition was measured with a pre-test and post-test administered before and after students engaged in the inquiry learning task.

Study 1

The first study investigated if providing partial hypotheses for students can support their hypothesis generation and facilitate their inquiry learning. By
providing partial hypotheses, we expected to ease the process of hypothesis generation and to direct students’ attention to more important and productive parts of the learning task. This study also explored if three different types of prior knowledge (knowledge about the inquiry process, general domain knowledge, and specific domain knowledge) influence the effect of the given hypothesis-generation support. Participants in this study, coming from the Netherlands, were provided with either terms representing the three main elements of a testable hypothesis or the same terms plus a partial hypothesis consisting of half a sentence giving the start of a hypothesis. Results indicated that students who were provided with partial hypotheses generated more complex hypotheses, performed better at data collection, and acquired more domain knowledge than students who were provided with terms only. No moderating effect of the three types of prior knowledge was found.

**Study 2**

Considering that the ready-made partial hypotheses are directive and to a certain extent limit students from expressing their own ideas in the hypotheses they complete, the support provided in Study 2 was less directive, aiming to prepare students to generate their own hypotheses. Inspired by previous literature indicating that students generate their hypotheses based on either their prior knowledge or their experiences from some prior exploratory experiments, we regarded providing relevant domain knowledge and exploratory practice as two possible ways to help students with their hypothesis generation. Hence, in Study 2, we investigated if providing domain information together with exploratory practice or providing either of them separately can improve students’ hypothesis generation, subsequent inquiry processes and knowledge acquisition. Participants, coming from the UK, were provided with either domain information and an exploratory opportunity in the simulation-based representation of the learning domain, just the same domain information as in the first condition, just the same exploratory opportunity as in the first condition, or no additional support at all. Results showed that only students who were provided domain information alone knew more about the variables in the domain before generating hypotheses than those who received no additional support. There were no differences between conditions on the quality of hypothesis generation, subsequent inquiry processes, or knowledge acquisition. Providing domain information alone was the only condition in which students improved from pre- to post-test.
Study 3
Given that students differ in their levels of prior domain knowledge, providing domain information adapted to their prior knowledge was the support investigated in Study 3. Participants, coming from Italy and Romania, were given either adaptive domain information based on their prior domain knowledge and an aligned hypothesis scratchpad with terms on variables, relations, and conditions or no introduction to the domain and an aligned hypothesis scratchpad with terms on relations and conditions but no terms on variables. The provision of adaptive domain information was based on two steps. First, students’ prior knowledge was recorded with a selection tool, within which both relevant and irrelevant variables were provided. Students were asked to select variables that they thought were important to take into consideration to solve the given inquiry problem. Students were also asked to link the variables that they thought were related. Second, the selected variables and indicated links between variables were used as codes to determine the appropriate level of domain information (full introduction, simple introduction, or no further introduction) to be provided, which was automatically given to students before hypothesis generation. Due to COVID-19, we only managed to involve a limited number of students in the experiment, and a condition in which students were provided with non-adaptive domain information could not be included. Results showed that students who were provided with adaptive domain information together with terms on variables in their hypothesis scratchpad considered a greater variety of variables in their hypotheses, indicated more testable relations in their hypotheses, and wrote hypotheses with a higher level of informativeness regarding variables, relations and conditions. But surprisingly, the better performance on hypothesis generation did not result in better knowledge acquisition.

Conclusions
To sum up, providing students with partial hypotheses and giving students adaptive domain information (together with an aligned hypothesis scratchpad) were two effective interventions for improving students’ performance in generating hypotheses. Providing partial hypotheses may also foster students’ knowledge acquisition. The findings of the present dissertation imply that providing partial hypotheses, which steer students to informative hypotheses
while allowing them to bring in their own ideas when completing the incomplete parts, might be the appropriate level of support for secondary school students with relatively low prior knowledge of the domain and limited or no experience in inquiry learning. The quest to find what type and amount of support is good for whom continues.
Nederlandse samenvatting
Inleiding

Onderzoekend leren in natuurwetenschappelijke domeinen stelt leerlingen in staat om op conceptueel niveau kennis te ontwikkelen. Door hen actief te laten exploreren en ontdekken heeft onderzoekend leren de potentie om leerlingen zowel kennis over natuurwetenschappelijke domeinen als kennis van wetenschappelijke methoden te laten verwerven. Dit kan echter niet worden bereikt door leerlingen zelfstandig in een onderzoekende leeromgeving te laten leren. Leerlingen kunnen moeilijkheden ondervinden in alle fasen van onderzoekend leren, zoals het opstellen van hypothesen, het ontwerpen van experimenten, dataverzameling en -interpretatie en het trekken van conclusies. Om te profiteren van onderzoekend leren, hebben leerlingen ondersteuning nodig die hen begeleidt bij de bijbehorende complexe processen.

In dit proefschrift richten wij ons op één van de processen die deel uitmaken van onderzoekend leren — het opstellen van hypothesen. Door leerlingen hypothesen te laten opstellen wordt een oppervlakkige verwerking van informatie tijdens een onderzoekend leren activiteit voorkomen en wordt hun cognitieve betrokkenheid bij het proces van onderzoekend leren gestimuleerd. Het opstellen van hypothesen vraagt leerlingen om bewust na te denken over het betreffende probleem, het probleem te relateren aan wat ze al weten, en hun eerste ideeën te structureren en om te zetten naar verbanden tussen variabelen. Daarnaast kunnen de opgestelde hypothesen richting geven bij het opzetten en uitvoeren van experimenten en bij het trekken van conclusies. Leerlingen ondervinden echter vaak moeilijkheden bij het opstellen van hypothesen. Volgens de literatuur weten sommige leerlingen niet hoe een hypothese die getoetst kan worden eruit ziet en vinden ze het moeilijk om hypothesen op te stellen en om mee te beginnen, hebben sommige leerlingen moeite met het bepalen van relevante variabelen en slagen sommigen er niet in om verbanden tussen variabelen te leggen.

Het doel van dit proefschrift was daarom mogelijke manieren te onderzoeken om leerlingen te ondersteunen bij het opstellen van hypothesen tijdens het onderzoekend leren. Daarbij moet de juiste mate van ondersteuning aangeboden worden. Het aanbieden van te veel ondersteuning beperkt de vrijheid van leerlingen om zelf beslissingen te nemen, terwijl het aanbieden van te weinig ondersteuning leerlingen belast met taken die (nog) te veeleisend voor hen zijn. De drie experimentele studies die deel uitmaken van dit proefschrift hadden daarom
als doel te bepalen welk type en hoeveel ondersteuning bij het opstellen van hypotheses het beste werkt voor leerlingen.

**Overzicht van de studies**

Alle experimentele studies in dit proefschrift werden uitgevoerd met middelbare scholieren (11-14 jaar) en bestonden uit drie sessies die plaatsvonden tijdens reguliere natuurkundelessen. De online onderzoekende leeromgevingen zoals die gebruikt werden in alle studies, zijn ontworpen met het Go-lab ecosysteem (https://www.golabz.eu) en behandelde hetzelfde natuurkundedomein – kracht en beweging. Dit domein is gekozen in overleg met de natuurkundedocenten van de deelnemende scholen om zo goed aan te sluiten bij hun curriculum. In alle gevallen hebben we erop toegezien dat het onderwerp nog niet eerder aan bod was gekomen tijdens de lessen. In de basis was de structuur van de leeromgevingen en het op te lossen probleem dat daarin centraal stond voor alle studies gelijk. De exacte taken als onderdeel van het onderzoekend leerproces verschillen echter per studie, afhankelijk van de specifieke onderzoeksvraag.

In alle studies en condities stelden leerlingen hypotheses op met behulp van een hypothesetool: het hypotheseekladblok. In de standaardversie van het hypotheseekladblok worden de drie belangrijkste elementen van een toetsbare hypothese (variabelen, verbanden en voorwaarden) aangeboden als vooraf gedefinieerde vorm. Leerlingen kunnen de aangeboden termen verslepen om hun eigen hypothese samen te stellen. Daarnaast is er een "zelf invullen" optie beschikbaar, zodat de leerlingen ook hun eigen ideeën konden toetsen. Afhankelijk van het type ondersteuning dat in de betreffende studie werd geboden, waren er minimale verschillen in de exacte configuratie van het hypotheseekladblok.

Het effect van de aangeboden ondersteuning werd in alle studies bepaald aan de hand van de mate van kennisverwerving van leerlingen en de kwaliteit van hun onderzoekend leerproces, in het bijzonder de kwaliteit van de opgestelde hypotheses. Aangezien de verbanden tussen variabelen in hypotheses richtinggevend zijn voor de daaropvolgende onderzoekend leerprocessen, inclusief dataverzameling en het trekken van conclusies, onderzochten we niet alleen het effect van de aangeboden ondersteuning op het opstellen van hypotheses, maar ook op de daaropvolgende onderzoekend leerprocessen. Om de kwaliteit van de onderzoekend leerprocessen van de leerlingen te bepalen zijn
Nederlandse samenvatting

codeerschema’s ontwikkeld. Kennisverwerving werd gemeten met behulp van een voor- en natoets, die werden afgenomen voordat en nadat leerlingen werkten aan de onderzoekend leertaak.

Studie 1

In de eerste studie is onderzocht of het aanbieden van gedeeltelijke hypothesen het opstellen van hypothesen kan ondersteunen en het onderzoekend leren van leerlingen kan faciliteren. Met het aanbieden van gedeeltelijke hypothesen verwachten we het proces van het opstellen van hypothesen te vergemakkelijken en de aandacht van leerlingen te richten op relevante onderdelen van de leertaak. Daarnaast werd in deze studie onderzocht of drie verschillende soorten voorkennis (kennis over het onderzoeksproces, algemene domeinkennis, en specifieke domeinkennis) het effect van de aangeboden ondersteuning ten aanzien van het opstellen van hypothesen zou beïnvloeden. Deelnemers aan deze studie, afkomstig uit Nederland, kregen ofwel termen aangeboden die de drie hoofdelementen van een toetsbare hypothese representeerden, ofwel diezelfde termen met daarbij een gedeeltelijke hypothese bestaande uit een halve zin waarbij het begin van een hypothese werd weergegeven. Resultaten toonden aan dat leerlingen die gedeeltelijke hypothesen aangeboden kregen, complexere hypothesen opstelden, beter presteerden tijdens de dataverzameling, en meer domeinkennis verwierven dan leerlingen die enkel de begrippen aangeboden kregen. Er werd geen modererend effect gevonden van de drie soorten voorkennis.

Studie 2

Aangezien vooraf opgestelde gedeeltelijke hypothesen sturend zijn en ze leerlingen in zekere mate beperken in het tot uiting brengen van hun eigen ideeën voor hypothesen, was de aangeboden ondersteuning in Studie 2 minder directief. Het idee was dat dit de leerlingen zou kunnen helpen om hun eigen hypothesen op te stellen. Afgaande op de literatuur waarin beschreven wordt dat leerlingen hun hypothesen opstellen op basis van ofwel hun voorkennis, ofwel hun ervaringen opgedaan tijdens eerdere explorerende experimenten, beschouwden we het aanbieden van relevante domeingereleaveerde informatie en exploratie van het domein als twee potentiële manieren om leerlingen te helpen bij het opstellen van hun hypothesen. In Studie 2 hebben we daarom onderzocht of het aanbieden van domeingereleaveerde informatie in combinatie met exploratie, of van één van beide afzonderlijk, kan leiden tot verbetering in het opstellen van hypothesen, de
Nederlandse samenvatting

daaropvolgende onderzoekend leerprocessen en de kennisverwerving van leerlingen. Deelnemers, afkomstig uit het Verenigd Koninkrijk, kregen ofwel domeingeralerde informatie en de mogelijkheid om een simulatie van het kennisdomein te exploreren, ofwel enkel dezelfde domeingeralerde informatie als in de eerste conditie, ofwel enkel dezelfde mogelijkheid om te experimenteren als in de eerste conditie, ofwel helemaal geen aanvullende ondersteuning. Resultaten toonden aan dat het enkel aanbieden van domeingeralerde informatie leerlingen hielp hun kennis over de betreffende variabelen te vergroten alvorens ze hypothesen opstelden in vergelijking met leerlingen uit de controle conditie. De andere twee condities scoorden niet beter op kennis van variabelen dan de controle conditie. Er waren geen verschillen tussen condities op de kwaliteit van opgestelde hypothesen, de daarop volgende onderzoekend leerprocessen en kennis acquisitie. De conditie waarin enkel domeingeralerde informatie aangeboden werd was de enige conditie waarin leerlingen een verbetering lieten zien van voor- naar natoets.

Studie 3

Omdat leerlingen kunnen verschillen in het niveau van voorkennis over het domein, werd in Studie 3 het effect onderzocht van het aanbieden van domeingeralerde informatie aangepast aan de voorkennis van de leerlingen. Deelnemers, afkomstig uit Italië en Roemenië, kregen ofwel adaptieve domeingeralerde informatie gebaseerd op hun voorkennis samen met een daarop afgestemd hypothesekladblok met termen omtrent variabelen, verbanden en voorwaarden aangeboden, ofwel geen domeingeralerde informatie maar wel een afgestemd hypothesekladblok met termen omtrent verbanden en voorwaarden, zonder begrippen omtrent variabelen, aangeboden. Het aanbod van de adaptieve domeingeralerde informatie was gebaseerd op twee stappen. Ten eerste werd de voorkennis van de leerlingen vastgesteld met een selectietool, waarin zowel relevante als irrelevant variabelen werden aangeboden. Leerlingen werd gevraagd variabelen te selecteren die zij belangrijk achten om in overweging te nemen bij het oplossen van het gegeven onderzoekend leerprobleem. Daarna werd hen gevraagd om de variabelen die volgens hen met elkaar verband hielden aan elkaar te linken. Ten tweede werden de geselecteerde variabelen en de aangegeven verbanden tussen variabelen gebruikt om het passende niveau van de adaptieve domeingeralerde informatie te bepalen (uitgebreide informatie, basale informatie, of geen verdere informatie), die automatisch aan de leerlingen werd aangeboden alvorens hen gevraagd werd hypothesen op te stellen. Vanwege COVID-19 kon slechts een beperkt aantal leerlingen aan het onderzoek deelnemen.
Nederlandse samenvatting

en kon een conditie waarin leerlingen niet-adaptieve domeingerelateerde informatie ontvingen niet worden gerealiseerd. Resultaten toonden aan dat leerlingen die adaptieve domeingerelateerde informatie ontvingen in combinatie met variabelen termen in hun hypothesekladblok een grotere variëteit aan variabelen in overweging namen in hun hypothesen, meer toetsbare verbanden in hun hypothesen aanduiden en informatiever hypothesen opstelden voor wat betreft variabelen, verbanden en voorwaarden. Opvallend was dat betere prestaties voor wat betreft het opstellen van hypothesen niet resulteerden in betere kennisverwerving.

Conclusie

Samenvattend kan worden gesteld dat het aanbieden van gedeeltelijke hypothesen en adaptieve domeingerelateerde informatie (in combinatie met een daarop afgestemd hypothesekladblok) twee effectieve interventies bleken om de prestaties van leerlingen bij het opstellen van hypothesen te verbeteren. Het aanbieden van gedeeltelijke hypothesen bevorderde ook de kennisverwerving van leerlingen. De bevindingen van dit proefschrift impliceren dat het aanbieden van gedeeltelijke hypothesen, die leerlingen wijzen op informatieve hypothesen terwijl het hen toestaat hun eigen ideeën in te brengen bij het aanvullen van de incomplete delen, een passend ondersteuningsniveau zou kunnen zijn voor de doelgroep bestaande uit middelbare scholieren (leerlingen met relatief weinig voorkennis van het domein en minimale of geen ervaring met het opstellen van hypothesen en onderzoekend leren). De zoektocht naar welk type en hoeveel ondersteuning er moet worden gegeven aan welke leerlingen gaat door.
Appendices
## Appendix A

Coding scheme for inquiry processes of Study 1

<table>
<thead>
<tr>
<th>Inquiry process</th>
<th>Data source</th>
<th>Coding item</th>
<th>Explanation</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hypothesis</td>
<td>Hypothesis</td>
<td>Testability</td>
<td>Valid independent variable 0- no independent variable is mentioned</td>
<td>1</td>
</tr>
<tr>
<td>generation</td>
<td>scratchpad</td>
<td></td>
<td>1- resultant force</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2- left force and right force</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3- motion (describing the motion of an object)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>4- speed</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>5- force</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>6- others</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>If a variable on force or motion is stated in the first half hypothesis</td>
<td></td>
</tr>
<tr>
<td>Valid condition</td>
<td></td>
<td></td>
<td>If a measurable or observable condition of the independent variable</td>
<td>1</td>
</tr>
<tr>
<td>of independent</td>
<td></td>
<td></td>
<td>is stated in the first half hypothesis</td>
<td></td>
</tr>
<tr>
<td>variable</td>
<td></td>
<td></td>
<td>Valid dependent variable 0- no dependent variable is mentioned</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1- resultant force</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2- left force and right force</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3- motion (describing the motion of an object)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>4- speed</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>5- force</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>6- others</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>If the dependent variable is stated and is not the same as the independent</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>variable in the second half hypothesis</td>
<td></td>
</tr>
<tr>
<td>Valid outcome</td>
<td></td>
<td></td>
<td>If a measurable or observable outcome condition is stated for the dependent</td>
<td>1</td>
</tr>
<tr>
<td>condition of</td>
<td></td>
<td></td>
<td>variable</td>
<td></td>
</tr>
<tr>
<td>the dependent</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>variable</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Complexity of the hypothesis</td>
<td>Variable selection</td>
<td>If the hypothesis stated the relationship between force and motion rather than only either of them</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>-------------------------------</td>
<td>--------------------</td>
<td>------------------------------------------------------------------------------------------------</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td>Variable focus</td>
<td>If the hypothesis focused on a balanced situation of force</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data collection</td>
<td>Data recording table</td>
<td>Data on the initial state of motion</td>
<td>If the initial state of motion of the object is recorded as stated in the hypothesis</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Data on the independent variable</td>
<td>If the recorded condition of the independent variable is in line with the condition stated in the hypothesis</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Data on the outcome condition of the dependent variable</td>
<td>If the condition of the stated dependent variable is recorded</td>
<td>1</td>
</tr>
<tr>
<td>Drawing conclusions</td>
<td>Input box</td>
<td>Conclusion about the hypothesis</td>
<td>If the conclusion is right accept the hypothesis that should be accepted reject the hypothesis that should be rejected</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Input box + data recording table</td>
<td>Data-based conclusion</td>
<td>If the student’s conclusion can be inferred from the recorded data</td>
<td>1</td>
</tr>
<tr>
<td>Reflection (final reflection)</td>
<td>Input box</td>
<td>Force can change motion</td>
<td>If the student generally mentioned that force can change the state of motion of an object</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>If the student clarified when net force ≠0 (or other similar expression), the state of motion of an object will change</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Input box</td>
<td>Motion can be maintained with force</td>
<td>If the student generally mentioned that force can maintain the state of motion of an object</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>If the student mentioned when net force = 0, or forces are balanced (or another similar expression), the force can maintain the state of motion of an object (either keep moving or remain stationary)</td>
<td>2</td>
</tr>
</tbody>
</table>
Appendices

Appendix B

Questions and coding scheme for responses on variable-related knowledge test

- Question 1: What is a force?
- Question 2: When we describe a force, what are the main characteristics to be mentioned?
- Question 3: What is friction?
- Question 4: What is the motion of an object?
- Question 5: When we describe the motion of an object, what are the main characteristics to be mentioned?
- Question 6: If you are asked to find out how forces affect the motion of an object, which variables (and their characteristics) do you need to take into consideration?

<table>
<thead>
<tr>
<th>Coding items</th>
<th>Score</th>
<th>Explanations</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Definition of force</td>
<td>0</td>
<td>No response or irrelevant response</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>Partial description of force as either push or pull</td>
<td>Pull OR Push</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Description of force as energy or strength</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>Both push and pull mentioned</td>
<td>Pull AND Push</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Description of force as energy or strength, as an attribute of physical action or movement</td>
<td></td>
</tr>
<tr>
<td>Variables on force</td>
<td>0</td>
<td>No response or irrelevant response</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>Only one sub-variable for force mentioned</td>
<td>Magnitude (size)(strength) OR Direction of application of force</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>Both sub-variables for force mentioned</td>
<td>Magnitude (size)(strength) AND Direction of application of force</td>
</tr>
<tr>
<td>Coding items</td>
<td>Score</td>
<td>Explanations</td>
<td>Examples</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>-------</td>
<td>---------------------------------------------------</td>
<td>---------------------------------------------------------------</td>
</tr>
<tr>
<td>Definition of friction</td>
<td>0</td>
<td>No response or irrelevant response</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>Description of a phenomenon associated with the concept</td>
<td>Force that prevents objects from moving</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>Scientific explanation of the concept</td>
<td>Force that always acts in the opposite direction of motion</td>
</tr>
<tr>
<td>Definition of motion</td>
<td>0</td>
<td>No response or irrelevant response</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>Description of a phenomenon associated with the concept</td>
<td>The movement of an object</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Something is moving or being moved</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>Scientific explanation of the concept</td>
<td>A change in the position of an object</td>
</tr>
<tr>
<td>Variables on motion</td>
<td>0</td>
<td>No response or irrelevant response</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>Only one sub-variable for motion mentioned</td>
<td>Speed OR direction of motion</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>Both sub-variables for motion mentioned</td>
<td>Speed AND direction of motion</td>
</tr>
<tr>
<td>Variables to take into</td>
<td>0</td>
<td>No response or irrelevant response</td>
<td></td>
</tr>
<tr>
<td>consideration</td>
<td>1-6</td>
<td>Number of variables on the variable list mentioned</td>
<td>1. Magnitude (size) of a force</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2. Direction of application of force</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3. Friction</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>4. Initial state of motion of an object</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>5. Speed</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>6. Direction of motion</td>
</tr>
</tbody>
</table>
# Appendix C

## Coding scheme for inquiry processes of Study 2 and 3

<table>
<thead>
<tr>
<th>Inquiry process</th>
<th>Coding variable</th>
<th>Coding item</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Hypothesis generation</strong></td>
<td>Number of hypotheses generated</td>
<td>Number of hypotheses generated</td>
<td>Count the number of hypotheses generated by a student.</td>
</tr>
<tr>
<td><strong>Diversity</strong></td>
<td>Diversity of the stated variables</td>
<td>Diversity of the stated variables</td>
<td>Count the number of different variables used in all testable hypotheses written by a student (count after checking the testability of the generated hypotheses, and count each variable only once if it is repeatedly stated).</td>
</tr>
<tr>
<td><strong>Testability</strong></td>
<td>Independent variable</td>
<td>What is the independent variable?</td>
<td>1. applied force</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2. friction</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3. sum of forces</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>4. force</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>5. direction of applied force</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>6. direction of friction</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>7. direction of sum of forces</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>8. direction of force</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>9. speed</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>10. direction of motion</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>11. general state of motion (e.g., moving or not)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>12. mass</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>13. other valid variable (e.g., size of the object)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>14. no valid variable</td>
</tr>
</tbody>
</table>
**Inquiry process** | **Coding variable** | **Coding item** | **Explanation** |
---|---|---|---|
Dependent variable | What is the dependent variable? | 1. applied force |  
|  | 2. friction |  
|  | 3. sum of forces |  
|  | 4. force |  
|  | 5. direction of applied force |  
|  | 6. direction of friction |  
|  | 7. direction of sum of forces |  
|  | 8. direction of force |  
|  | 9. speed |  
|  | 10. direction of motion |  
|  | 11. general state of motion (e.g., moving or not) |  
|  | 12. mass |  
|  | 13. other valid variable (e.g., size of the object) |  
|  | 14. no valid variable |  

If the variables stated in the hypothesis are on the topic of force and motion (from the list above), and the dependent variable is not the same as the independent variable, then this is regarded as a valid ‘variable set’ and 1 point will be given.

**Relation selected** | **If a condition of an independent variable is stated and a corresponding outcome condition of a dependent variable is also indicated, then the relation will be regarded as a valid relation and 1 point will be given.**

**Informativeness** | **The level of informativeness of the stated variables** |
---|---|
Informativeness of variables | Level 0: No response/non-testable hypothesis |
<table>
<thead>
<tr>
<th>Inquiry process</th>
<th>Coding variable</th>
<th>Coding item</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level 1: Both independent and dependent variables are sub-variables of the same general variable (both about force or both about motion).</td>
<td>Example: If the applied force is equal to the friction, then the sum of force is 0.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Level 2: Both forces and motion are stated, forces are components of force (applied force and/or friction), and only a general state of motion (will move or will not move) of an object is mentioned.</td>
<td>Example: If applied force is greater than friction, then the object will move.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Level 3: Both forces and motion are stated, forces are components of force (applied force and/or friction), and the characteristics of motion are clarified (speed or moving direction). Or the sum of forces is stated but only a general state of motion is mentioned.</td>
<td>Examples: If applied force is greater than friction, then the speed of the object will increase. If the sum of forces is not 0, then the object will move.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Level 4: Both forces and motion are stated, forces are focused on the sum of forces, and the characteristics of motion are clarified.</td>
<td>Example: If the sum of forces is 0, then the speed of the object will remain the same.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Informativeness of relation**

The level of informativeness of the relation between variables
<table>
<thead>
<tr>
<th>Inquiry process</th>
<th>Coding variable</th>
<th>Coding item</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Level 0: No response/non-testable hypothesis</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Level 1: The hypothesis is about a direct relation between two variables.</td>
<td>Example: If the applied force increases, then the friction increases.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Level 2: The hypothesis is about an indirect relation, via intermediate variables. (The sum of forces is coded as level 2. But if the variables are all about force, then the sum of forces is coded as level 1).</td>
<td>Example: If the applied force is greater than the friction, then the speed of the object will increase.</td>
</tr>
<tr>
<td></td>
<td>Informativeness of condition</td>
<td>If the initial state of motion is stated in the hypothesis, then one point will be given.</td>
<td></td>
</tr>
<tr>
<td>Data recording</td>
<td>Initial state of motion</td>
<td>If the initial state of motion of the object is recorded, then one point will be given.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Condition of the independent variable</td>
<td>If the condition of the independent variable is described, then one point will be given.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Outcome condition of the dependent variable</td>
<td>If the outcome condition of the dependent variable is described, then one point will be given.</td>
<td></td>
</tr>
<tr>
<td>Drawing conclusions</td>
<td>Conclusion about the hypothesis</td>
<td>If the conclusion is correct, then one point will be given. accept the hypothesis that should be accepted reject the hypothesis that should be rejected</td>
<td></td>
</tr>
<tr>
<td>Inquiry process</td>
<td>Coding variable</td>
<td>Coding item</td>
<td>Explanation</td>
</tr>
<tr>
<td>-----------------</td>
<td>-----------------</td>
<td>-------------</td>
<td>-------------</td>
</tr>
<tr>
<td></td>
<td>Data-based conclusion</td>
<td>If the student’s conclusion can be inferred from the recorded observations, then one point will be given.</td>
<td></td>
</tr>
</tbody>
</table>

**Final summary**

<table>
<thead>
<tr>
<th>Coding variable</th>
<th>Coding item</th>
</tr>
</thead>
<tbody>
<tr>
<td>Force can change motion</td>
<td>0 — No response/totally not relevant</td>
</tr>
<tr>
<td></td>
<td>1 — Generally mentions that force can make something move/stop (e.g., when applying force, an object can move)</td>
</tr>
<tr>
<td></td>
<td>2 — Correctly mentions some specific situations when force can change some characteristics of motion (e.g., when applied force is greater than friction, an object can move)</td>
</tr>
<tr>
<td></td>
<td>3 — Conclusively states that when sum of forces ≠ 0 or &gt; 0 (or other similar expression), the state of motion of an object will change</td>
</tr>
</tbody>
</table>

**Motion can be maintained with force**

<table>
<thead>
<tr>
<th>Coding variable</th>
<th>Coding item</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 — No response/totally not relevant</td>
<td></td>
</tr>
<tr>
<td>1 — Generally mentions that force can maintain the state of motion of an object (when there is no force)</td>
<td></td>
</tr>
<tr>
<td>2 — Correctly mentions some specific situations when force can maintain some characteristics of motion (remain stationary only when sum of forces = 0)</td>
<td></td>
</tr>
<tr>
<td>3 — States that when net force = 0 or forces are balanced (or another similar expression), the force can maintain the state of motion of an object (it will either keep moving or remain stationary)</td>
<td></td>
</tr>
</tbody>
</table>
Hypothesis generation, one of the major processes of inquiry learning, is a process to organize what you know about a domain to make sense of a problem and to organize tentative answers as testable propositions to guide further investigation. It is a good starting point for students to get cognitively engaged in the inquiry activity. However, students frequently experience difficulties in hypothesis generation. Hence, this dissertation aimed to investigate possible ways to support hypothesis generation by students during inquiry learning.

Finding the right level of support is a tricky balancing act. Three different types of support were designed, implemented, and investigated in the three studies involved in this dissertation, attempting to explore the provision of just-enough support for students’ hypothesis generation.

Results indicated that providing partial hypotheses might be the appropriate level of support for secondary school students with relatively low prior knowledge of the domain and limited or no experience in inquiry learning.