The Individual Cognitive Resources Underlying Students' Mathematical Argumentation and Proof Skills

From Theory to Intervention

Daniel Sommerhoff

München 2017

The Individual Cognitive Resources Underlying Students' Mathematical Argumentation and Proof Skills

From Theory to Intervention

Dissertation zum Erwerb des

Doctor of Philosophy (Ph.D.)

am Munich Center of the Learning Sciences

der Ludwig-Maximilians-Universität München

Vorgelegt von Daniel Sommerhoff

aus San Francisco

München, 4. Mai 2017

Credit Title Image: Marish/Shutterstock.com

Erstgutachter: Prof. Dr. Stefan Ufer (Ludwig-Maximilians-Universität München)
Zweitgutachter: Prof. Dr. Ingo Kollar (Universität Augsburg)

Weitere Mitglieder der Prüfungskommission:
Prof. Alan Schoenfeld (University of California, Berkeley)
Prof. Dr. Frank Fischer (Ludwig-Maximilians-Universität München)

Datum der mündlichen Prüfung: 12.07.2017

The Individual Cognitive Resources Underlying Students' Mathematical Argumentation and Proof Skills

From Theory to Intervention

Abstract

Handling mathematical argumentation and proof proficiently is regarded as an important learning goal within mathematics in general and at the beginning of university education in particular. At university, mathematical proof is introduced as the *central* method of mathematics as a scientific discipline. Despite this importance as a major learning goal, students were repeatedly shown to have severe difficulties with mathematical argumentation and proof. With the MIMAPS project (Measuring and Improving Mathematical Argumentation and Proof Skills) we strive to obtain insights into first-year university students' mathematical argumentation and proof skills and to develop and evaluate strategies to effectively support students in the acquisition of these skills. We were especially interested in the consequences of interpreting mathematical argumentation and proof skills as a complex cognitive skill that depends on several underlying individual cognitive resources such as knowledge facets and skills.

To structure our research, a framework by Blömeke, Gustafsson, and Shavelson (2015) for competencies and underlying resources was adapted to mathematical argumentation and proof skills taking prior mathematics education research into consideration. The resulting framework fosters a comprehensive view of mathematical argumentation and proof skills, combining three aspects, the underlying individual *resources*, the *situations* that require the use of mathematical argumentation and proof skills, and the *processes* leading to an observable performance.

As an initial step of the project a *descriptive literature review* was conducted. Results revealed that a comprehensive approach to mathematical argumentation and proof skills has rarely been taken in mathematics education research. Most studies rather focus on specific resources, processes, or situations in the context of mathematical argumentation and proof skills. In particular, many studies consider one or two resources of mathematical argumentation and proof skills but do not examine multiple resources at the same time, so that little evidence on the relative influence of individual resources on mathematical argumentation and proof skills exists.

Based on these findings, we conducted a *correlational study* assessing six individual cognitive resources of first-year mathematics students as well as their performance in proof construction and proof validation. Data were quantitatively analyzed using Generalized Linear Mixed Models to examine the relative influence of the individual resources on students' performance in proof construction and proof validation. Results verify and extend prior findings that mathematical argumentation and proof skills are knowledge intensive, that is participants' mathematical content knowledge has a substantial influence. Further, both domain-specific and domain-general resources have an influence on students' mathematical argumentation and proof skills, yet the domain-specific resources predominate. Also mathematical strategic knowledge, which has been little researched so far, plays an important role both in proof construction and in proof validation. Our study could not replicate the influence of problem-solving skills on mathematical argumentation and proof skills that were previously shown to have an impact in secondary school geometry proof construction contexts (Chinnappan, Ekanayake, & Brown, 2012; Ufer, Heinze, &

Reiss, 2008). Among several explanations, this difference may particularly be related to varying conceptualizations and operationalizations of problem-solving skills in these studies, and to the selection and use of domain-general *vs*. domain-specific strategies to approach mathematical proofs, which may change with the level of experience and mathematical expertise. Thus, pupils starting with mathematical proof may rely more on weak, domain-general problem-solving strategies, whereas university students may have more reliable, domain-specific strategies at hand.

In a third step, an *intervention study* was designed that explicitly acknowledges the resources underlying mathematical argumentation and proof skills. The intervention was intended to support the resources of mathematical argumentation and proof skills and thereby indirectly also the overall skill using two different approaches inspired by research from instructional design: The *one-by-one approach* explicitly separated the instruction regarding the four potential resources of mathematical argumentation and proof skills and supported them individually. The *simultaneous approach*, however, focused on the resources concurrently, creating an integrated learning experience and fostering connections between the resources. Results emphasize that students in both conditions improved substantially regarding most resources. Despite being antithetic, both approaches yield comparable learning effects, implying that the central tenet of the part-task / whole-task debate that whole-task learning is superior, cannot be directly transferred to the *resource-based* approaches for supporting students' mathematical argumentation and proof skills. The results for overall mathematical argumentation and proof skills are mixed, and initially weaker students appear to benefit more from the intervention, especially from the simultaneous approach.

In summary, our project revealed that mathematical argumentation and proof skills indeed depend on several individual cognitive resources but that their relative influence differs largely when analyzing their impact in a comprehensive way. The project also gave statistical evidence for the individual importance of each resource and underlined the impact of domain-specific resources. It further showed that resources can be supported effectively and that they give rise to novel ways of instruction on mathematical argumentation and proof skills. The MIMAPS project thus contributes to current research both with a research framework fostering a comprehensive perspective on mathematical argumentation and proof skills as well as by three studies providing several new findings related to research on students' individual underlying resources, their importance, and ways to support these. Several questions arose, particularly regarding the interplay and importance of problem-solving skills and mathematical strategic knowledge, which need to be addressed in future studies.

Keywords: Argumentation, Proof, University Students, Complex Cognitive Skills

Acknowledgements

The research presented in this work was supported by the Elite Network of Bavaria [Project number: K-GS-2012-209]. I would like to extend my sincere gratitude for the opportunities (conference attendance, incubator stay, international knowledge exchange) made possible by the ENB.

Preface

> Some policy-makers believe that supporting educational research is crazy, but surely the real madness is to carry on what we have been doing, and yet to expect different outcomes.
>
> (Black & Wiliam, 2003, p. 635)

As a student and tutor for mathematics at the LMU, I often felt that even those students who participated actively in tutorials would leave after 90 minutes still harboring problems and open questions. Discussing this issue with other tutors, lecture assistants, and professors, I realized that many had accepted this as a status quo. Furthermore, none of them were able to offer a satisfying approach to change the situation other than based on anecdotal evidence from their own teaching. Thus, I felt the need to change something in the way of teaching mathematics at university and to explore how students could be supported more effectively. Yet, at that time, I did not know how.

I want to express my sincere gratitude to my supervisors Prof. Dr. Stefan Ufer and Prof. Dr. Ingo Kollar for co-initiating the REASON program and thereby bringing me on the right track to contribute to this very goal and to join educational research. During the last years, both brought me up as a (junior) member of the scientific community. They did a terrific job in introducing me to the pathways of science, its nature, goals, and rules, as well as helping me to acquire the knowledge and skills needed to participate in it. Both gave me guidance where I needed it and freedom where I could handle it. And, of course, both had helpful advice guiding me in the right direction to improve my work. Thank you so much!

Especially during my stay at UC Berkeley my third supervisor, Prof. Alan Schoenfeld, greatly contributed to my work and my future in research. Intense, inherently too short discussions on the dialectic between theoretical, qualitative, and quantitative research helped me to see connections that were hidden before and to gain a more meaningful perspective by attending to all three aspects in my research.

Besides my supervisors, many other people at LMU helped me find my way in life, in research in general, and in my topic in particular. I want to thank Prof. Dr. Frank Fischer, coordinator of the REASON program, who gave me the chance to get behind-the-curtain views into academia and participate in the NAPLeS-Initiative. Furthermore, I want to thank the mathematics education group as well as the REASON-Team, especially Sarah Ottinger, Simon Weixler, Katharina Engelmann, Esther Beltermann, Christian Ghanem, and Sandra Becker for the great time, the discussions, and the chance to learn so much from and with them, be it for work or life.

Outside of LMU, especially the members of the mathematics education group from TUM, Esther Brunner, and the *functions* research group at UC Berkeley deserve my gratitude.

Finally, I want to thank my family, in particular my parents, my brother, and my girlfriend for supporting and encouraging me during the last years and for the frequent fruitful discussions we had.

Completing my dissertation, another step towards the future is achieved, and yet there is so much unknown to come. But being accompanied by such great colleagues, friends, and family, I am happy to take the next steps and embrace the future.

> Educational research can and does make a difference, but it will succeed only if we recognize its messy, contingent, fragile nature.
>
> (Black & Wiliam, 2003, p. 635)

Table of Content

List of Figures

List of Tables

1 Introduction

Scientific communities from different disciplines have distinct methods to generate scientific knowledge. Physicists conduct experiments, psychologists employ experimental methods such as observation or intervention studies, literature researchers apply the hermeneutic circle and mathematicians use proof. Since the ancient Greeks, proof is so inherent to the nature of mathematics as a science (Jahnke, 2010) and so unique to mathematics that mathematics is called a *proving science* (Heintz, 2000). Today, handling *mathematical argumentation* is an important goal within mathematics classrooms world-wide (e.g., Common Core State Standards Initiative, 2010; KMK, 2012; National Council of Teachers of Mathematics, 2000) and first proofs are introduced to students in school, for example, Bhaskara's proof of the Pythagorean theorem (see Nelsen, 1993, p. 4) or the proof for the irrationality of the square root of 2. Yet, the first time students get in touch with proof as a *scientific method* relying on formalism and an axiomatic underpinning is at university when they first encounter proof-based analysis or linear algebra lectures in their mathematics studies.

In Germany and many other countries, this introduction of proof based mathematics coincides with the transition from secondary school to university. Research from the past decades (e.g., M. Clark & Lovric, 2009; Grünwald, Kossow, Sauerbier, & Klymchuk, 2004; Moore, 1994; Rach, Heinze, & Ufer, 2014; Reichersdorfer, Ufer, Lindmeier, & Reiss, 2014; A. Selden, 2011; Tall, 1991) has repeatedly shown that students struggle exactly at this point and has given substantial evidence that their struggles are at least in part caused by problems with handling mathematical proof (e.g., A. Selden, 2011; A. Selden & Selden, 2013; Tall, 1992, 2008; Weber, 2003). Studies within the German educational system reflect students' difficulties quantitatively: After the first semester 34% - 45% of mathematics students drop out (Dieter & Törner, 2012) and approximately 47% of the German students starting a mathematics bachelor program leave university without any degree (Heublein, Richter, Schmelzer, & Sommer, 2014). Furthermore, data from the United States of America, for example by Seymour and Hewitt (1997), as well as preliminary data from the recent "Talking about leaving, revisited" project (Seymour & Ferrare, 2015) also indicate high switching rates (30 - 60%) from mathematics to other subjects.

One explanation for students' difficulties in handling mathematical argumentation and proof and the high interindividual variation in students' performance brought forward by prior research is the fact that mathematical argumentation and proof skills comprise a complex cognitive skill (e.g., Chinnappan et al., 2012; Reichersdorfer et al., 2014; Reiss, Heinze, & Klieme, 2002; Schoenfeld, 1985), in Europe often even conceptualized as a competence (e.g., Klieme & Leutner, 2006; Koeppen, Hartig, Klieme, & Leutner, 2008; Weinert, 1999). That is, mathematical argumentation and proof skills are conceptualized as the latent disposition to handle mathematical arguments and proofs, which requires the integration of several different resources such as knowledge facets, skills, or beliefs, and is enacted via several processes within different situations.

The assumption that several individual resources are needed to be successful in situations involving mathematical argumentation and proof skills, can be easily illustrated. It is impossible for a student to construct a mathematical proof without having the mathematical content knowledge at hand (e.g., the definitions of the objects within the task) or knowing the acceptance criteria a mathematical proof is subject to. In consequence, identifying those underlying resources and estimating their relative importance could offer valuable knowledge when trying to support students. Results would not only give further theoretical insights on the structure of mathematical argumentation and proof skills as a complex cognitive skill, but also uncover those resources required to be successful in handling mathematical argumentation and proof and thereby possibly disclose ways to support students in acquiring these skills.

The approach of examining the resources underlying mathematical argumentation and proof skills complements prior qualitative and quantitative research (e.g., Chinnappan et al., 2012; Schoenfeld, 1985; Ufer et al., 2008; Weber, 2001) and can be based on Schoenfeld's claim (2012a, p. 231):

> People's in-the-moment decision making when they teach, and when they engage in other well practiced, knowledge intensive activities, is a function of their knowledge and resources, goals, and beliefs and orientations. Their decisions and actions can be "captured" (explained and modeled) in detail using only these constructs.

Schoenfeld (1985, 2012a) gave evidence for this claim in various studies, focusing on problem solving and mathematical proof construction as well as on a set of non-mathematics related activities. Similar approaches were also taken by other researchers (e.g., Chinnappan et al., 2012; Ufer et al., 2008) and for other topics and areas such as teaching skills or vocational education (e.g., Mulder, Gulikers, Biemans, & Wesselink, 2009; Shulman, 1987).

The fact that Schoenfeld's claim is about "knowledge intensive activities" in general and not only mathematical argumentation and proof skills or problem-solving skills, highlights another issue mainly unattended by mathematics education research, that is the relation of mathematical argumentation and proof skills to *scientific reasoning and argumentation skills* in general (e.g., Dunbar & Klahr, 2012; Fischer, Kollar, et al., 2014; Klahr & Dunbar, 1988; Kuhn, 2002; Zimmerman, 2000). Although proof is a genuine mathematical method, the processes used for handling mathematical argumentation and proof, such as generating hypotheses or new evidence, and the demands posed in these situations are closely related to domain-general scientific reasoning and argumentation processes, which in this case handle mathematical objects and are subject to the norms of the local mathematical community. Embedding mathematical argumentation and proof skills in this way into the broader research context allows to connect to and possibly utilize prior research from psychology, education, and related fields. Here, especially research on the processes of scientific reasoning and argumentation may be useful to describe mathematical argumentation and proof skills, leading to a shared terminology and insights that can be compared and possibly transferred within different domains in future.

Even though students' struggles with mathematical proof have been recognized by research for a long time and several attempts have been made to support students (e.g., Mevarech & Fridkin, 2006; Moore, 1994), up to now the situation hardly changed little as drop out and switching rates are still comparatively high (Dieter & Törner, 2012; Heublein et al., 2014). Realizing these shortcomings, we initiated the MIMAPS project (*Measuring and Improving Mathematical Argumentation and Proof Skills*) to create seminal knowledge regarding theory as well as practice. Here, we report results from three studies that shed light on students' mathematical argumentation and proof skills, the activities of proof construction and proof validation, as well as their relation to diverse content-specific, domain-specific, and domain-general individual cognitive resources. These results are then used as a basis for two instructional approaches to support students' mathematical argumentation and proof skills.

In the spirit of Mark Twain's quote

> It ain't what you don't know that gets you into trouble. It's what you know for sure that just ain't so.

the project responds to existing qualitative studies (e.g., Schoenfeld, 1985; Schoenfeld, 2012a; Weber, 2001) and tries to quantitatively investigate the generalizability of prior research findings regarding the resources of mathematical argumentation and proof skills in the context of university mathematics.

As basis for this project we created a research framework based on the work by Blömeke et al. (2015), conceptualizing mathematical argumentation and proof skills as a complex cognitive skill, enclosing not only the underlying resources, but also the situations that require mathematical argumentation skills as well as the processes leading to an observable performance. The framework was gradually refined throughout the project to match prior research findings from mathematics education research as well as those within the project.

In a first step, we conducted a literature review on research focusing on mathematical argumentation and proof skills in secondary and tertiary contexts. Results highlight current research foci regarding several resources, processes, as well as situations in the context of mathematical argumentation and proof and how these are examined and combined within research.

Consequently, in a correlational study, we assessed six cognitive resources suggested by prior research as well as students' performance in two situations that require their mathematical argumentation and proof skills. Using Generalized Linear Mixed Models, we closely examined the relative impact of the individual resources in the context of *proof construction* – the creation of a mathematical proof for a given claim – and *proof validation* – the reading and judging of the correctness of a given mathematical proof. Here, three domain-specific and three domain-general individual cognitive resources that were suggested by prior research were compared regarding their relative influence in both situations. Results extend prior findings that handling mathematical argumentation and proof is a knowledge-intense activity, yet not all potential resources could be validated as underlying mathematical argumentation and proof skills.

The third study was designed as an intervention study, trying to support students in their acquisition and learning of overall mathematical argumentation and proof skills by means of four suggested underlying resources. Two different approaches based on the part-task / whole-task debate from instructional design (e.g., Branch & Merrill, 2011; van Merriënboer & Kester, 2007), focusing on the resources *one-by-one vs.* focusing on them *simultaneously*, were compared. Analyses contrast both approaches regarding their effects on the resources as well as on overall mathematical argumentation and proof skills and examine their feasibility.

These three studies are presented in chapter 5, preceded by three theoretical chapters. First, the basis of this thesis is formed by trying to disentangle the notions of reasoning, argumentation, and proof and highlighting the social character of mathematical proof. Furthermore, a bridge towards scientific reasoning and argumentation is built (chapter 2) to put the research within this project into a larger frame and enable a critical discussion of the applicability of a framework for scientific reasoning and argumentation skills in the context of mathematical argumentation and proof. Building on this foundation, the current state of research on mathematical argumentation and proof skills is discussed (chapter 3). Here, an existing general framework for complex cognitive skills, covering their various aspects is introduced and the different aspects of the framework are discussed in depth in the context of mathematical argumentation and proof thereafter. We elaborate on the various resources suggested by prior research as underlying mathematical argumentation and proof skills, various frameworks for processes involved when handling mathematical argumentation and proof, as well as various situations, respectively activities in this context. Both chapters are fundamental for chapter 4 displaying our research framework, which describes mathematical argumentation and proof skills in a comprehensive way, incorporating various resources, processes, and situations based on prior research findings, and describing the research questions guiding this project.

The final chapter 6 comprehensively presents the results from all three studies, highlights central aspects, and discusses them in an integrated way. Finally, the limitations of the project and its findings are discussed, and an outlook containing implications and a research agenda with important questions that could not be addressed or arose in the current studies is given.

2 Reasoning, Argumentation, and Proof

Outline *The terms reasoning, argumentation, and proof are widely used, yet there are no generally accepted definitions. To form a solid theoretical base for this thesis, we first review definitions of reasoning, argumentation, and proof in the context of diverse theoretical frameworks and research traditions. Afterwards, we outline the inherently social nature of mathematical proof and the relevance of socio-mathematical norms for students' argumentation and proof skills. Finally, we position mathematical argumentation and proof skills in the more general research on scientific reasoning and argumentation to highlight connections between both and to address further facets and underlying research that are often not recognized within the scope of mathematics education research.*

2.1 Delineating and Separating Terms

Although often mentioned together, *reasoning*, *argumentation*, and *proof* refer to at least partially distinct activities and associated products. All three are part of a relatively large cluster of terms, also including for example *justification*, *proving* and *arguing*, which are defined vaguely and partially overlapping in (mathematics) education research (e.g., Aberdein, 2009; Cirillo, Kosko, Newton, Staples, & Weber, 2015; Osborne & Patterson, 2011; Reid & Knipping, 2010). Their definitions, meanings, and connotations differ according to research traditions (or perspectives), the particular focus of research, and the specific data being analyzed. Thus, they are still subject to some debate and confusion (e.g., Reid & Knipping, 2010). The framework for this project outlined below is based on terms and definitions from (educational) psychology as well as mathematics education to be able to connect to research from both disciplines. Additionally, some alternative definitions are shortly described to highlight differences and to avoid misconceptions.

Within the three terms reasoning, argumentation, and proof, *reasoning* is often interpreted as the most elementary activity, defined as the process of drawing a conclusion from principles and evidence (Leighton & Sternberg, 2003). The result of this activity can be either a statement, thought or decision (often implicit in action), which has to have some systematic relation to the given principles, evidence or more generally speaking premises. The most common example in the context of mathematics is drawing a deductive inference, but also inductive or abductive reasoning as well as reference to authority fall within this definition (Johnson-Laird, 2000). The term reasoning, therefore, is broad, yet limited to single steps of inference rather than multiple.

Although the term *argument* can be interpreted as a single reason for or against something, a structured sequence of reasons and claims, or as a certain type of social exchange (Hornikx & Hahn, 2012), the term *argumentation* within educational psychology and mathematics education is often used in the second sense as a sequence of inferences or assertions, leading from a given premise to a conclusion (e.g., Douek, 2007; Halpern, 2002; Reiss & Ufer, 2009; Toulmin, 2003). Contrasting reasoning and argumentation, this way of defining argumentation implies that reasoning can be seen as drawing single inferences, whereas argumentation rather refers to the combination of multiple inferences to a (more or less) logical chain in order to provide evidence for or against a given claim (Reichersdorfer et al., 2012; Reiss, Heinze, Renkl, & Groß, 2008) or explore a given task, problem, or situation (e.g., Reiss & Ufer, 2009). Following these definitions, reasoning is obviously needed for argumentation and can be seen as an underlying activity.

In contrast, other research traditions use the term reasoning for what we defined as argumentation (e.g., Lithner, 2008). The difference between reasoning and argumentation is then often attributed to a shift between the aims underlying these activities or their setting: In this tradition, reasoning is usually used to denote an individual activity (e.g., reasoning about a phenomenon when reading about it), whereas argumentation is used to denote reasoning processes within groups (e.g., Andriessen, 2009; Felton & Kuhn, 2001; Voss & Van Dyke, 2001), often also explicitly aimed at convincing group members or others that a claim, hypothesis, or point of view is correct or true (e.g., van Eemeren & Grootendorst, 1999).

Here, we will follow the first definition of argumentation. Based on it, argumentation can have various forms and features, for example, it can be more or less long, deductive, logically correct, and / or convincing. Each of these characteristics is of varying importance for its acceptance, depending on the context and its aim and can be used to evaluate a given argumentation (see section 2.2). This fact is often used to define a *(mathematical) proof*[1] as a certain kind of argumentation that satisfies certain norms within the mathematical community. One such norm, which is often mentioned, is the sole acceptance of deductive inferences (see further Dawkins & Weber, 2016; Heinze & Reiss, 2003). For example, Meyer (2007, p. 21) defines constructing a mathematical proof as an activity where a claim is formally, step-by-step deductively inferred from known theorems or definitions in a valid way. In comparison to argumentation, three characteristics of proof stand out: "formally", "deductively", and "inferred from known theorems or definitions in a valid way". All three features refer to socio-mathematical norms (see Yackel & Cobb, 1996) that define what is required for an argumentation to be a proof. Requiring all three, Meyer's (2007) definition is close to what is often called a formal proof (Hales, 2008; Hanna & Jahnke, 1993), thereby relatively strict and close to what is sometimes called derivation (Aberdein, 2009). Hales (2008, p. 1371) defines a (completely) formal proof as

> a proof in which every logical inference has been checked all the way back to the fundamental axioms of mathematics. All the intermediate logical steps are supplied, without exception.

This definition of proof corresponds to an *ideal proof* that is hardly ever achieved in mathematical practice as checking inferences back to the fundamental axioms or providing all logical steps would make proofs incredibly long and mainly useless for practice (e.g., Hanna, 1989; Jahnke & Ufer, 2015). Accordingly, proofs are usually constructed and communicated in an abbreviated way where "routine logical steps are omitted" (Hales, 2008, p. 1371) in such a way that the proof remains *in principle formalizable* (Alama & Kahle, 2013). A definition taking this into account is provided by Bell (1976, p. 26), defining proof very openly as

> a directed tree of statements, connected by implications, whose end point is the conclusion and whose starting points are either in the data or are generally agreed facts or principles

and mainly resembles our definition of argumentation.

In this project we decided to adhere to an intermediate course taken by A. J. Stylianides (2007), defining a proof as a mathematical argument, that is a connected sequence of assertions, that is subject to several characteristics[2] and social norms. His definition corresponds well to the recent suggestion by Weber (see Cirillo et al., 2016) that

---

[1] Within this project, we conceptualize proof mainly as an object and examine student's skills to handle these as well as the processes involved in handling them. Thus, we do not discuss the *concept of proof* in detail, which would be an additional topic (see further Reid & Knipping, 2010).

[2] In his article, he mentions three categories of these characteristics: *set of accepted statements*, *modes of argumentation*, and *modes of argument representation.*

> proof might be best understood as a cluster concept as in the case of Lakoff (1990). The idea is that proof in mathematical practice might be an argument that contains a large number of features (e.g., being purely deductive, highly convincing, perspicuous, within a representation system, and socially sanctioned) but does not necessarily contain all of them with no single feature being common amongst every proof. Consequently, there is not a decision procedure in which you can definitively say that an argument is (or is not) a proof, if it has (or lacks) some property and one should avoid elevating any singular property (e.g., being convincing) as the essence of proof (see Weber, 2014, for a discussion of this issue).

Except for these restrictions in form and features for an argument to become a proof, a distinction between the epistemological viewpoint as well as a logical and cognitive distance is often highlighted (Balacheff, 1988; Boero, Garuti, Lemut, & Mariotti, 1996; Duval, 1995; Pedemonte, 2007). (Pedemonte, 2008, p. 385) speaks of a

> "structural gap" between argumentation and proof because in argumentation inferences are based on content while in proof they follow a deductive scheme (data, claim, and inference rules).

Epp (2003) denotes this gap as a shift towards a "different logical and linguistic world" than the one "normally inhabited" and Downs and Mamona-Downs (2005) compare learning mathematical proof with learning an "entirely new language". This linguistic aspect is also supported by J. Selden and Selden (1995), who point out that producing mathematically sound statements (i.e., proofs or parts of it) out of informally written statements (i.e., arguments) is a significant barrier when constructing and validating proofs.
Both difference and connection between argumentation and proof are reflected in Brunner's (2013) continuum between argumentation as an everyday kind of reasoning[3], which in her framework is the umbrella term for argumentation and proof, and formal-deductive mathematical proof (Figure 1). Between both ends of the spectrum is a continuum of more or less mathematics-related and more or less deductive kinds of argumentation, for example, logical argumentation underpinned with mathematical formulas. The connection between argumentation and proof is also emphasized by the introduction of *cognitive unity* by several Italian researchers (e.g., Boero, Garuti, & Mariotti, 1996; Mariotti, 2006; Pedemonte, 2007) who despite of the structural gap see a productive connection between argumentation and proof and highlight the benefits of argumentation for proof.

[3] The German term used by Brunner is "Begründen".

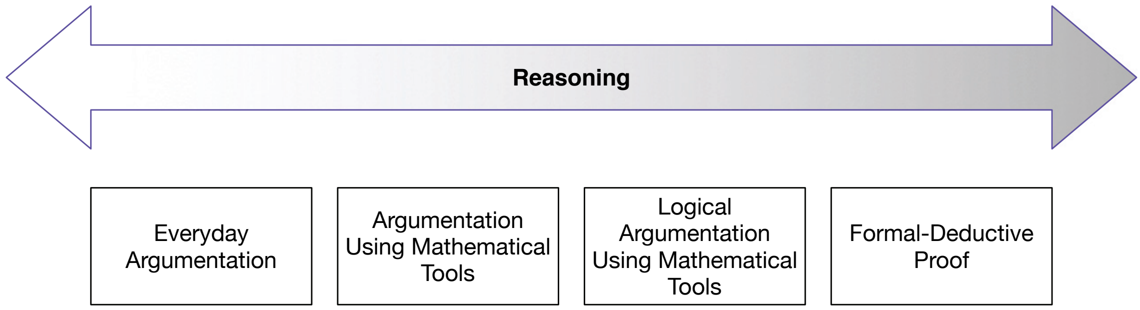


Figure 1. The continuum of reasoning connecting argumentation and proof (Brunner, 2013).

The perspective on argumentation and proof introduced in this thesis corresponds best to the tradition of *proving as problem solving* (G. J. Stylianides, Stylianides, & Weber, in press), because we implicitly view the creation of an argumentation and more specifically of a proof as a special case of problem solving where students are asked to use deductions (or other inferences) to achieve the goal of proving a claim and thereby solving the given problem. The other two traditions (*proving as convincing* and *proving as a socially embedded activity*) mentioned by G. J. Stylianides et al. (in press) are of minor importance here, although we highly acknowledge both aspects. We also deem conviction (of oneself and / or others) as an important aim of handling mathematical proofs (i.e., for both proof construction and proof comprehension). Yet, we do not limit handling mathematical proofs or more generally argumentation to this goal as multiple researchers have outlined a broader set of goals for proofs (e.g., Bell, 1976; de Villiers, 1990; Hanna, 1990). Furthermore, as discussed above, the concept of mathematical proof is dependent on socio-mathematical norms, which presupposes a social embedding.

2.2 Mathematical Proof in the Context of Communities

As introduced above, mathematical proofs are argumentations that are subject to certain norms, so-called *socio-mathematical norms* (e.g., Yackel & Cobb, 1996), which qualify them to be accepted as mathematical proofs. These norms are not intrinsic to mathematical proof but are established by the local mathematical community. A famous quote from Manin (2010, p. 45) puts it this way:

> A proof becomes a proof only after the social act of "accepting it as a proof".

Devlin (2003) illustrates this by showing that even today a "proof", which may be 100% correct, still must be socially accepted to be "correct". For this he uses Hales' purported proof (2002) of Kepler's sphere-packing conjecture. It was rejected by the mathematical community after four years of close examination by twelve referees because they were not able to ascertain its correctness.

Hence, mathematical proofs and doing mathematics, in general, cannot be seen as an absolute concept, but must always be regarded within the social context (e.g., Op 't Eynde, Corte, & Verschaffel, 2002), because working with proofs is "an inherently social activity" (Schoenfeld, 1992, p. 335). Crucial for this perspective is the fact that there is not *one* mathematical community, but that there are multiple communities and that the concept of proof is dependent on shared norms within these. For example, in some 7th grade classrooms, the act of ripping off two vertices, respectively angles, of a paper triangle and placing these next to the third angle, so that a straight line is generated (Figure 2), which students already know corresponds to a straight angle of 180°, may be accepted as a proof for the sum of interior angles of a triangle.

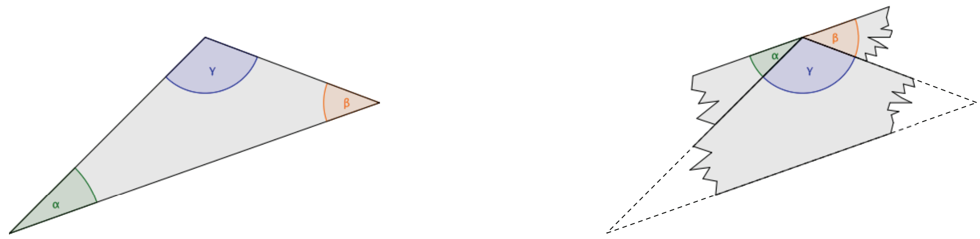


Figure 2. Sketch of the "proof" for the sum of interior angles by ripping of the vertices of a triangle.

In other classrooms, this "proof" may not be accepted. Here a proof using the parallel postulate and alternate angles may be constructed (Figure 3), thus explicitly referring to already known mathematical objects and properties. Research mathematicians, on the other hand, may require even more information, for example about the underlying geometrical space, to be sure that the parallel postulate as the argument connecting the given data and claim (i.e., *warrant* using Toulmin's (2003) terminology) holds.

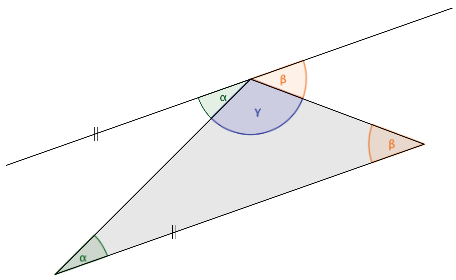


Figure 3. Sketch of the proof using the parallel postulate.

A further issue in the context of the social character of mathematical proofs, especially within education, is the diversity of (didactical) proof concepts. Already in his first sentence, Bell (1976, p. 23) speaks of proofs as the aspect of mathematics that probably "shows the widest variation in approaches". Wittmann and Müller (1988) broadly distinguish between three different types of (didactical) proofs

- formal-deductive proof
- experimental proof
- operative proof ("inhaltlich-anschaulicher Beweis")

These types of proofs are dissected even further by Brunner (2013, p. 52 ff), using a framework on proof approaches by Leiss and Blum (2006), and also by Biehler and Kempen (2016). Which of these didactical proof concepts are accepted as a valid mathematical proof and which do not (independent of the quality of the individual purported proof) is again up to the context and community. Shifts regarding the acceptance of these proof concepts in different communities can be the cause of serious issues and problems, for example for students in the transition towards proof-based mathematics courses within university. Tall (1992, p. 495 ff) describes this transition as a shift from concept, intuition, and experience based mathematics towards a kind of mathematics based on formal definitions and logical deductions.

Transitioning from school to university, the local mathematical community changes, other norms apply, and hence different proof concepts are accepted (e.g., Biehler & Kempen, 2016; Dreyfus, Nardi, & Leikin, 2012; Kempen, 2016). For example, operative proofs may have been very well accepted and even encouraged at school, but may not suffice at university. It is therefore mandatory for first-year students to recognize and learn the norms applied at university in general and specific courses in particular.

Weber, Inglis, and Mejía-Ramos (2014, p. 11) further propose aligning university students' proof concepts with those of (research) mathematicians as a goal of university mathematics instruction. This process of alignment has been researched by scientists from several backgrounds, including mathematics education, psychology, sociology, and anthropology. Descriptions range from a rather one-sided *enculturation* of the new students in the already existing university community by communicating these norms, to social-constructivist views of shared co-construction of norms by lecturers, tutors, and students (e.g, Hemmi, 2008; Müller-Hill & Kempen, in preparation; Nickerson & Rasmussen, 2009; Pfeiffer, 2011). For example, Perry, Samper, Camargo, Molina, and Echeverry (2009) highlight that learning to prove corresponds to the participation in proving activities within the community of mathematical discourse, following the idea of entering a community of practice (Wenger, 1999), rather than to a "passive" enculturation.

This issue is aggravated by the fact that the transition into the university mathematical community neither happens instantly nor that there is one "mathematical community" (not even in mathematical research), but that there may be several communities for example in different content or research areas. Finally, there is increasing evidence that the norms used in teaching at university are not equivalent to those of research mathematicians (e.g., Weber et al., 2014).

2.3 Mathematical Argumentation and Proof Skills

The two past sections highlight that mathematical proofs can be seen as a subset of mathematical argumentation and that the distinction between both relies on socio-mathematical norms. Further, there is a continuum (Brunner, 2013) between both argumentation and proof (see Figure 1) that partially reflects the enculturation of students to the concept of mathematical proof by applying increasingly strict norms regarding their validity. For first-year university students, this enculturation has not reached its final state, which is for example reflected by findings that teaching practices of mathematicians do not reflect their research practices (e.g., Heintz, 2000; Hersh, 1999; Weber et al., 2014).

Based on this notion and understanding of proof, we will therefore in the following use the term *mathematical argumentation and proof skills* to denote university students' skills to handle argumentations according to the norms of the university mathematical community, for example, to construct valid proofs or to validate purported proofs against these norms. We thereby consider the fact that the acceptance criteria for argumentations to be "mathematical proofs" at the university are not identical to the standards of the international mathematical research community, which are aimed at mathematical proofs by researchers. Thus students' purported proofs are often positioned between what we above called argumentation and mathematical proof. This approach can be considered as analogue to the term reasoning-and-proving introduced by G. J. Stylianides (2008) to account for the various formulations used for proof and argumentation related activities in school curricula.

2.4 Connecting Mathematical Proof to Scientific Reasoning

The concept of *proof* is genuine to mathematics and proof is often characterized as "the most important characteristic of modern mathematics" (Hanna, 1991). Yet, other disciplines also engage in similar scientific activities to ascertain claims, and ultimately argumentation can be found throughout research and more practice related activities in all disciplines. Thus, considering research findings on argumentation skills from other disciplines could be highly valuable for research on mathematical argumentation and proof skills to connect and contrast findings in particular as underlying psychological processes and phenomena can be expected to be closely related.

One way of interlinking and setting mathematical argumentation and proof skills into a broader frame can be achieved using the concept of *scientific reasoning skills* (e.g., Bao et al., 2009; Klahr & Dunbar, 1988; Kuhn, 2002), which currently is referred to as a 21st century skill (e.g., National Research Council, 2012; OECD, 2013) and is promoted by educational initiatives globally (e.g., Obama, 2009). Although the term "reasoning" is used, the term *scientific reasoning* is not correctly interpreted when using the definition of reasoning outlined in section 2.1. Here, reasoning is meant in a broader sense, which is why scientific reasoning is also often referred to as *argumentative thinking*, *scientific thinking* (see Zimmerman, 2000), or lately also as *scientific reasoning and argumentation* (SRA)[4] (Fischer, Kollar, et al., 2014), attempting to show that it embraces more than drawing single inferences.

As Opitz, Heene, and Fischer (2015) point out, three different conceptualizations of scientific reasoning and argumentation are common in literature (Figure 4): Two of them conceptualize scientific reasoning and argumentation mainly on a skill-level. In accordance with Inhelder and Piaget (1958), one conceptualization views scientific reasoning as one general ability. The other conceptualization, for example, taken by Livermore (1964), is characterized by viewing scientific reasoning and argumentation as a collection of several, basically unrelated skills. Finally, the third conceptualization focuses more on the process of scientific reasoning, which is regarded in close relation to the process of problem solving. The latter perspective is taken for example by Klahr and Dunbar (1988). Generally, Opitz et al. (2015) observed a shift towards this last conceptualization within the assessment of scientific reasoning and argumentation skills in the last 30 years.

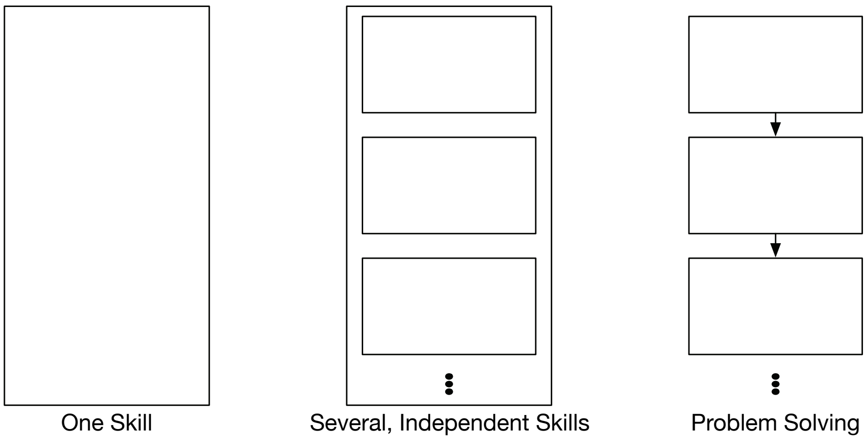


Figure 4. Different conceptualizations of scientific reasoning (Opitz et al., 2015).

According to Schunn and Anderson (1999), the term *scientific reasoning* covers all skills that are required for making scientific discoveries and progress. Thereby scientific reasoning "involves a diverse collection of cognitive activities, rather than single cognitive process" (Schunn & Anderson, 1999, p. 337). Mathematical argumentation and proof skills therefore easily fit in this

[4] Throughout this thesis the terms "scientific reasoning" and "scientific reasoning and argumentation" will be used mostly equivalently, as the latter solely highlights the inclusion of argumentation within reasoning but is otherwise equivalent to "scientific reasoning".

line of research as mathematical proof is a central scientific method within the field of mathematics to make scientific progress and as it also comprises several activities (see section 3.3). To proceed like this has several advantages:

- A cross-disciplinary language, which renders research results understandable as well as comparable between domains, is used.
- It allows the application of process models of scientific reasoning and argumentation in the context of mathematical argumentation and proof skills.
- Existing research results from domains other than mathematics, especially from psychology and education, can be taken into account when doing research on mathematical argumentation and proof skills.
- It allows to differentiate between domain-specific and domain-general aspects of mathematical argumentation and proof skills.
- It simplifies the process of finding commonalities between different domains, for example, to find cross-domain or domain-general aspects of (mathematical) argumentation and proof (see further Fischer, Wecker, et al., 2014).
- It allows researchers to combine parts of this project's data with data from other studies on scientific reasoning and argumentation when meta-analytically reanalyzing the data.

Research on scientific reasoning originates in the 1950s as a consequence of the cognitive revolution, with groundbreaking work by Bruner, Goodnow, and Austin (1956) as well as Inhelder and Piaget (1958). Here, seminal work was done within developmental psychology, later also by educational and more generally by cognitive psychology as well as science education. Since the 1950s, substantial research has been devoted to scientific reasoning (e.g., Dunbar & Klahr, 2012; Fischer, Kollar, et al., 2014; Giere, Bickle, & Mauldin, 2005; Klahr & Dunbar, 1988; Koslowski, 1996; Kuhn, 2002; Schunn & Anderson, 1999; Sodian & Bullock, 2008; Zimmerman, 2000) and several theoretical, as well as empirical advances, have been made.

2.4.1 The Dual Search Space

A major step forward in understanding and describing processes of scientific reasoning and also mathematical argumentation and proof was made by Klahr and Dunbar (1988) by creating the *dual search model of scientific discovery*, which according to Klahr and Dunbar (1988, p. 32) is "a general model of scientific reasoning that can be applied to any context in which hypotheses are proposed and data is collected". The framework connects two different positions in research on scientific reasoning (Zimmerman, 2000, p. 101), as it encompasses domain-specific knowledge and domain-general knowledge, respectively strategies, and therefore allows to examine the interactions between both facets (see further van Joolingen & de Jong, 1997). It frames scientific reasoning as a process of problem solving, which can be described as a guided search, information gathering, and information processing task (see Newell & Simon, 1972; Simon, 1978; Zimmerman, 2000) that takes place in two separate but related search spaces (Klahr & Dunbar, 1988; van Joolingen & de Jong, 1997). Both *hypothesis space* and *experiment space* should be understood according to the information processing approach by Newell and Simon (1972) as representations of the problem situation that include an initial state, a goal state, and a set of operators that allow to advance from one state within that space to another state.

The crucial idea underlying the dual problem space approach is the existence of two separated spaces: The *hypothesis space* that is related to the search for a hypothesis, its formulation, and its refinement and the *experiment space* that is related to any kind of experimentation, data gathering, or evidence generation used to evaluate this hypothesis. Still, according to Klahr and Dunbar (1988) both are connected in such a way that the search in the hypothesis space is guided by prior knowledge and results from the experiment space, whereas search in the experiment space is influenced by the current hypothesis and available domain-general strategies (Figure 5).

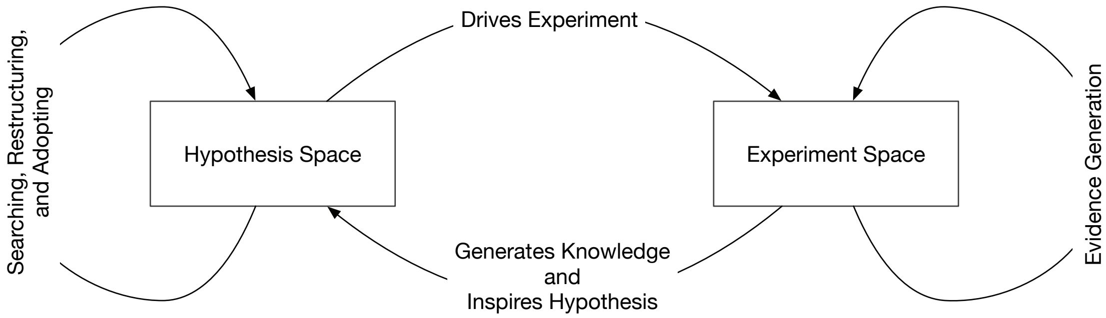


Figure 5. Operations and connections within the dual search space (Klahr & Dunbar, 1988).

The use of this framework by Klahr and Dunbar (1988) for mathematical argumentation and proof processes is not straight forward as it was intended for scientific domains that are based on mainly inductive empirical evidence, that is *data, statistics, artefacts, or materials that corroborate a claim or position*. Accordingly, the role of deductions within the framework and whether they should be regarded as part of the hypothesis space or the experiment space remains open. If they are attributed to the former, most processes in the context of mathematical proof are related to the hypothesis space, and the experiment space is limited to activities such as exploring and conjecturing as well as to illustrative objects such as specific examples or sketches. This approach has been taken for example by Philipp (2012), focusing explicitly on activities in the context of exploring. However, in the context of mathematical proof that mainly relies on deductive inferences, this approach appears less fruitful. If, in contrast, deductions are attributed to the experiment space and thereby are interpreted as a way to create mathematical evidence, the dual space framework can be used more purposefully for proof construction activities. The process of adapting a hypothesis however also requires deductive reasoning processes but is part of the hypothesis space. Accordingly, mapping deductions entirely to either the hypothesis space or the experiment space appear unreasonable in the context of mathematical proof. Thus, to purposefully use the framework in this context, deductions related to the hypothesis should be included in the hypothesis space whereas those related to creating mathematical evidence should be included in the experiment space.
An illustration of the framework by Klahr and Dunbar (1988) for mathematical argumentation and proof processes using this last approach is given in the example below. It is based on the following task given by Koedinger (1998, p. 320) and encompasses a newly-created description of a fictional proof construction process by a fictional student called Luke.

> A “kite” is a special kind of quadrilateral whose four sides form two pairs of congruent adjacent segments. In other words, a kite is a quadrilateral ABCD with AB congruent to CB and AD congruent to CD.
>
> Investigate these figures called kites using whatever tools you would like and discover and write down what must be true of every kite.

The description of the process is inspired by Koedinger’s (1998) task analysis and interview results. The student’s work is idealized to a certain degree, because his intentions and mental processes would of course not be accessible and the student may have made more faults or given up. Also, some “prolonging” steps have been omitted when they were not helpful for the demonstration of the dual search space.

Illustration of the use of the dual search space framework for mathematical argumentation and proof processes

Description	Coding
First thing after reading the task, Luke starts to draw a sketch of a kite. Since he has never done this before, he is not quite sure how to do this, and his first sketch ends up being just a regular quadrilateral and no kite (Figure 6). Being unhappy with the kite, he tries again, this time trying to "construct" one. He finally manages to do so by using several subsidiary circles (Figure 7).	Experiment space

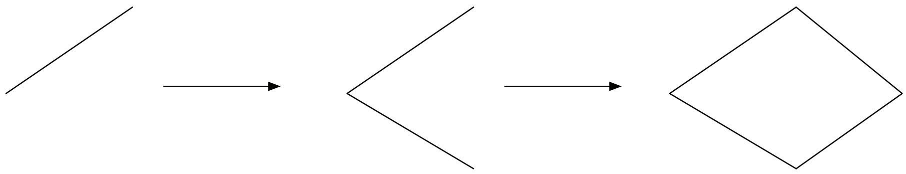


Figure 6. Creation of a first sketch of a kite.

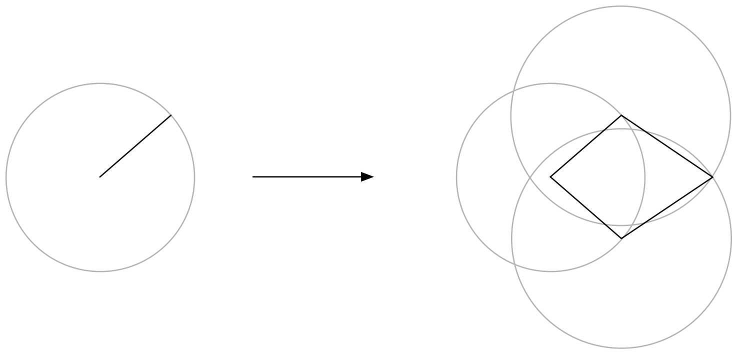


Figure 7. Creation of the first constructed kite.

Having created the kite, he carefully looks at it and forms the hypothesis that opposite angles are equal.	Hypothesis space
To confirm his hypothesis, he constructs another kite with arbitrary, but different lengths (Figure 8). By observing the angles, he sees two seemingly differing angles on the left and right of the kite. Carefully measuring the two seemingly different angles with his protractor Luke finds out,	Experiment space
that his initial hypothesis is invalid.	Hypothesis space

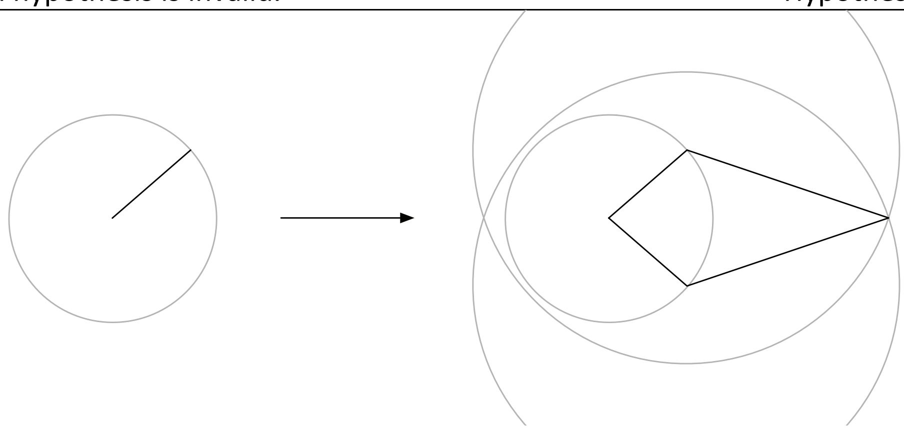


Figure 8. Creation of the second constructed kite.

Description	**Coding**
Not exactly knowing how to go on, he also measures the other two angles to see, whether they are also unequal. To his surprise, both angles are roughly the same size.	Experiment space
To confirm this, he measures the angles of the initial kite again, constructs another kite, and measures the angles of interest. For all three kites, both angles have the same size.	Experiment space
After doing so, he develops a new hypothesis: The angles at the top and bottom of his kites are of equal size.	Hypothesis space
To prove his hypothesis, he creates a new, bigger sketch (Figure 9), including both angles (highlighted at the points A and C), and some subsidiary lines. Enticed by the sketch, he has the initial idea to use the sum of interior angles and neighboring angles to show that both angles are equal. As he starts to compute $\angle$ DBA, he writes $\angle DBA + \angle BAC + 90° = 180°$ and therefore $\angle DBA = 180° - 90° - \angle BAC = 90° - \angle BAC$. Having written that down he pauses for a moment, reconsidering what he just wrote, because he had the feeling something was odd. After rereading what he had just written and looking at the sketch, he crosses out both lines because he realized that he does not know for sure that $\angle AMB = 90°$.	Experiment space

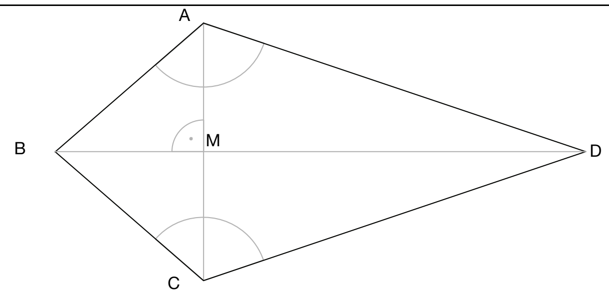


Figure 9. Kite with angles.

Having failed to use the angles, he looks at the sketch again, this time focusing more on "what he really knows". That basically amounts to $\overline{AB} = \overline{BC}$ and $\overline{AD} = \overline{DC}$. That in mind, he realizes that both "upper and lower" triangle have the diagonal BD in common, as well as two sides of the same length. He mumbles: Aha, that's how it works. It's true!	Experiment space
He, therefore, writes down: The triangle ABD and BCD have three sides of the same length. Per definition $\overline{AB} = \overline{BC}$ and $\overline{AD} = \overline{DC}$. The third line BD is literally the same in both triangles, so it has the same length. Knowing this, we can use the "side-side-side" rule to see that both triangles are congruent.	Experiment space
But then, as congruent triangles have the same angles, the angels $\angle BAD$ and $\angle DCB$ must be the same because they are the corresponding angles in both triangles.	Experiment space
Rereading what he has written, Luke is satisfied and concludes that his hypothesis was correct and is now proven.	Hypothesis space

2.4.2 Domain-specificity *vs.* Domain-generality

Examining the example above more closely, not only focusing on both spaces, it becomes clear that Luke benefits from his conceptual and procedural mathematical knowledge (e.g., de Jong & Ferguson-Hessler, 1996; Hiebert & Lefevre, 1986). For example, he uses his procedural mathematical knowledge on how to use the compass for his constructions, and conceptual knowledge such as the sum of interior angles or the "side-side-side" rule. Still, he may eventually have been able to construct the proof even without this underlying mathematical knowledge, that is only with domain-general knowledge and strategies (Abel, 2003; Chinnappan et al., 2012), such as drawing a sketch (which he did at the very beginning) and using, for example, thought experiments, argumentation (without referring to mathematical content knowledge), or folding.
The question about the domain-specificity or domain-generality of scientific reasoning and argumentation skills was a controversy from its very beginning. According to Osborne (see his contribution to Fischer, Wecker, et al., 2014) many definitions and frameworks for scientific reasoning tried to emphasize its domain-generality, that is they tried to point out that the skill is independent of domain-specific knowledge. This is also reflected by the review by Voss, Wiley, and Carretero (1995) who see scientific reasoning as a "general intellectual skill" and therefore as rather domain-general. This view is understood when focusing particularly on the processes involved in scientific reasoning activities, the underlying strategies, or the personal epistemological beliefs, rather than on the outcomes of scientific reasoning activities. For example, creating a hypothesis is a skill (or process, depending on the conceptualization of scientific reasoning), which (at least at first sight) appears as it can be applied across various domains. Yet, studies (e.g., Voss, Tyler, Yengo, & others, 1983) showed that there are significant group differences in the quality of the created hypotheses, with experts having domain-specific knowledge outperforming all other groups. Furthermore, experts from other domains, that is without domain-knowledge, performed not only significantly worse than the domain experts but also roughly on the level of undergraduate students, which were neither "experts" nor did they have extensive domain-specific knowledge.
Research from the last decades provided increasing evidence that scientific reasoning indeed is dependent on domain knowledge (e.g., Klahr & Dunbar, 1988; van Joolingen & de Jong, 1997; Voss et al., 1983), a view that today is mainly agreed upon by experts (Zimmerman, 2007). For example, a study by Schunn and Anderson (1999) with experts and students with diverse backgrounds such as psychology, physical science, social science, and arts revealed that domain-specific as well as domain-general resources can be beneficial in scientific reasoning and argumentation activities. In their study, the overall quality of domain-experts' solutions was superior to those of the other participants. However, they were also able to show that *domain-general experiment design skills* improved the quality of solutions.
Although today scientific reasoning and argumentation skills are acknowledged to be domain-specific to a certain degree, that is they draw on domain-specific knowledge facets such as conceptual, procedural, or strategic knowledge (see Kuhn, Schauble, & Garcia-Mila, 1992; Weber, 2001), the discussion is not finished (Fischer, Chinn, Engelmann, & Osborne, in preparation). Moreover, theoretical and educational questions regarding the transfer of strategies and the effectiveness of domain-general interventions in particular are still under debate as they are questioned by some researchers (e.g., Sweller, 1990; Tricot & Sweller, 2014) and emphasized by others (e.g., Chinnappan & Lawson, 1996; D. W. Eccles & Feltovich, 2008).
A reasonable approach, consistent with the findings regarding the importance of domain-specific knowledge, is to conceptualize scientific reasoning and argumentation skills as a complex cognitive skill that has several underlying skills and knowledge facets, which are partly domain-specific and partly domain-general. Still, the questions about the impact of each underlying

resource, the possibility of transfer between domains (of each resource and the overall skill), as well as ways to intervene (again, regarding each resource and the overall skill) remain widely open for future research.
Accordingly, scientific reasoning skills correspond to mathematical argumentation and proof skills regarding multiple aspects, for example regarding the underlying resources and their domain-specificity or -generality, which will be addressed in this project. Further, similarities can also be found on a process level.

2.4.3 The SRA-Framework

A recent approach to scientific reasoning and argumentation by Fischer, Kollar, et al. (2014) avoids presupposing any claim about domain-specificity or domain-generality or underlying resources by making an implicit definition of scientific reasoning and argumentation via three *epistemic modes* and eight *epistemic activities*, which describe the underlying aims and the processes of scientific reasoning and argumentation. Thereby, this approach belongs to the third category in the classification by Opitz et al. (2015) and describes scientific reasoning and argumentation skills as problem solving via observable processes.

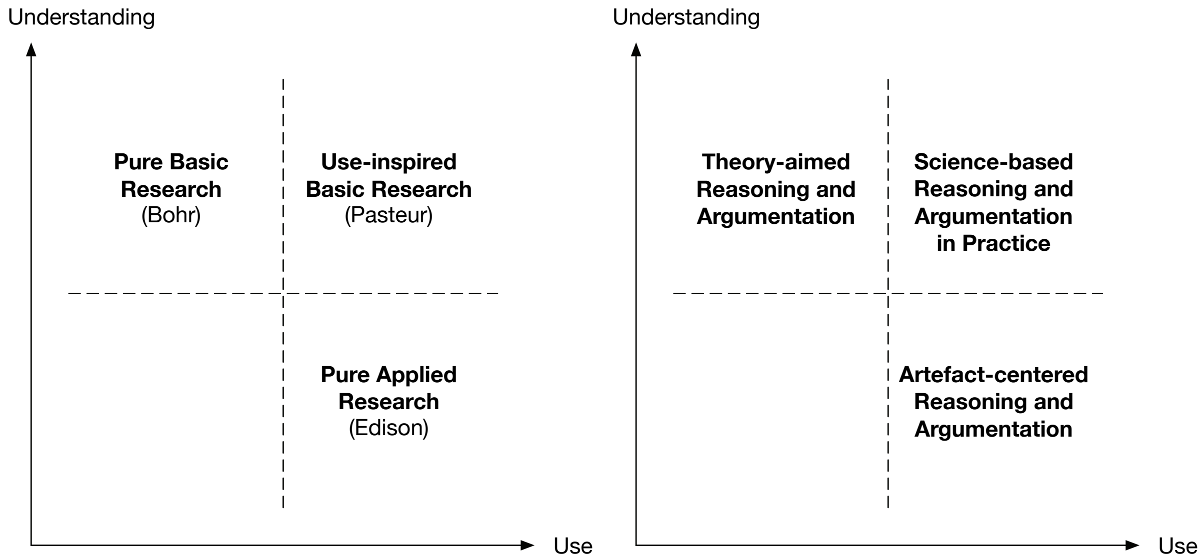


Figure 10. Epistemic modes of research (Stokes, 1997) (left) and scientific reasoning and argumentation (Fischer, Kollar, et al., 2014) (right).

To structure the context scientific reasoning and argumentation processes may occur in, Fischer, Kollar, et al. (2014) introduce three *epistemic modes* focusing on the aims underlying scientific reasoning and argumentation processes(Figure 10, right). These are based on work by Stokes (1997) (Figure 10, left) and describe three fundamentally different areas of scientific reasoning and argumentation based on two aims: *use* and *understanding*.

Theory-aimed reasoning and argumentation:
Scientific reasoning and argumentation activities mainly driven by the aim *understanding*, that is the wish to advance theory building about natural or social phenomena, are contained in the first epistemic mode, which is called *theory-aimed reasoning and argumentation* (Fischer, Kollar, et al., 2014). It is the analogue to Bohr's quadrant in Stokes' (1997) framework and is characterized by the generation and testing of hypotheses to explore underlying mechanisms, without having any application in mind. An example could be a student using experiments during a physics lesson to determine the local *g*-constant to understand the concept of gravity better.

Artefact-centered reasoning and argumentation:
The second epistemic mode, characterized by "use" as the underlying aim, is called *artefact-centered reasoning and argumentation* and is analogue to Edison's quadrant in Stokes' (1997) framework. It denotes scientific reasoning and argumentation that is centered around a specific artefact (e.g., a dynamic geometry software sheet in mathematics or a prototype in physics), often using circular, iterative steps of development of the artefact to solve a current problem. Although this is done using scientific methods and underlying theories, the activities are not aimed at creating a new theory that exceeds the one artefact / problem given.

Science-based reasoning and argumentation in practice:
The third epistemic mode of scientific reasoning and argumentation, analogue to Pasteur's quadrant, is labeled *science-based reasoning and argumentation in practice*. It denotes learners' scientific reasoning processes in practice, which are aimed at solving specific, real-world problems by building on scientific methods and theories and at the same time refining those. Most importantly, a solution of a given problem is neither solely academic in nature nor purely applied, but comprises both a practical and theoretical advancement. Typical, broad examples from mathematics could be financial mathematics research done in or on behalf of insurance companies. It is evoked by practical problems of the company, but ideally answers both, the practical problem as well as theoretical questions regarding probabilistic processes. Although associated closely with academia, many cases of design-based research (e.g., Collins, Joseph, & Bielaczyc, 2004; The Design-Based Research Collective, 2003) also belong to this category, as ideally both a theoretical advancement and a designed object (e.g., learning environment) are the aims and results of this methodology.

Building on the epistemic modes and the context they provide for scientific reasoning and argumentation, the framework by Fischer, Kollar, et al. (2014) describes eight processes, so-called *epistemic activities* that are assumed to be sufficient to describe the processes during any kind of scientific reasoning in either of the epistemic modes and across domains (Figure 11).

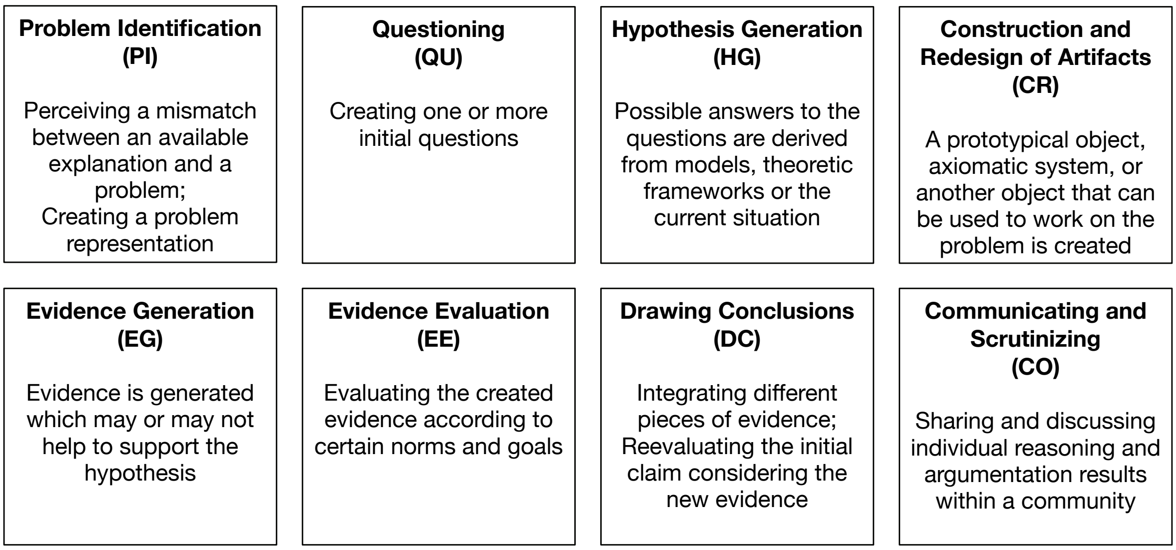


Figure 11. Epistemic activities of scientific reasoning and argumentation (Fischer, Kollar, et al., 2014).

It is worthwhile to notice that the epistemic activities do not have a specific order, do not have to be entailed in every scientific reasoning and argumentation process, and can be nested in several levels as well as repeated several times (Figure 12). A good example for this is a typical proof task given to university students, starting with “Prove, that X” (Iannone & Inglis, 2010; A. J. Stylianides & Stylianides, 2006, p. 205). The genteel reader will notice that a formulation like this stifles the activities of *questioning* and *hypothesis generation* because an (even not so experienced) student will instantly know that the claim (X) within the task is true and that the student “just” needs to find a way to prove it. Then, when actually proving the main claim (X) of the task by generating evidence for it and evaluating it, the student may encounter sub-problems (*problem identification*), which may lead to questioning, hypothesis generation, and so on (Figure 12). Also, when deducing one line of the proof after the other, the student may (and perhaps should (see Inglis & Alcock, 2012)) be in a constant transition between evidence generation (e.g., writing a line or deducting a fact) and evidence evaluation when checking the logical structure of that line and the consistency with what the student has done before.

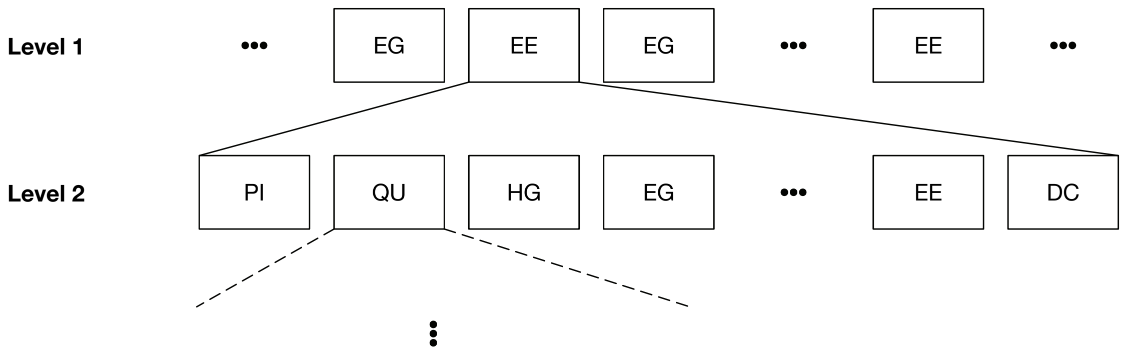


Figure 12. An example of epistemic activities nested in two levels during scientific reasoning and argumentation, with the second level being evoked during an evidence evaluation process.

To illustrate the use and application of the framework and in particular the epistemic activities in the context of mathematical proof construction processes, the example from Luke from the dual search space (see section 2.4.1) is picked up. Here, we describe the main steps on the first level of the task, partially giving short indications of second level (sub-problem) processes, without going into detail there.

Illustration of the use of the epistemic activities to code mathematical argumentation and proof processes

No	Description	Coding
1	First thing after reading the task,	Problem identification
2	Luke starts to draw a sketch of a kite. Since he has never done this before, he is not quite sure how to do this, and his first sketch ends up being just a regular quadrilateral and no kite (see Figure 6).	Construction of artefacts
3	Being unhappy with the kite, he tries again, this time trying to "construct" one. He finally manages to do so by using several subsidiary circles (see Figure 7).	Sub-problem: Evidence evaluation; Drawing Conclusions; Construction of artefacts
4	Having created the kite, he carefully looks at it and forms the hypothesis that opposite angles are equal.	Hypothesis generation
5	To confirm his hypothesis, he constructs another kite with arbitrary, but different lengths (see Figure 8).	Evidence generation
6	By observing the angles, he sees two seemingly differing angles on the left and right of the kite.	Evidence evaluation
7	Carefully measuring the two seemingly different angles with his protractor Luke finds out,	Evidence generation
8	that his initial hypothesis is invalid.	Evidence evaluation
9	Not exactly knowing how to go on, he also measures the other two angles to see, whether they are also unequal.	Evidence generation
10	To his surprise, both angles are roughly the same size.	Hypothesis generation
11	To confirm this, he measures the angles of the initial kite again, constructs another kite, and measures the angles of interest.	Evidence generation
12	For all three kites, both angles have the same size.	Evidence evaluation
13	After doing so, he develops a new hypothesis: The angles at the top and bottom of his kites are of equal size.	Hypothesis generation
14	To prove his hypothesis, he creates a new, bigger sketch (see Figure 9), including both angles (highlighted at the points A and C), and some subsidiary lines.	Construction of artefacts

No	Description	Coding
15	Enticed by the sketch, he has the initial idea to use the sum of interior angles and neighboring angles to show that both angles are equal.	Evidence generation (or sub-problem)
16	As he starts to compute $\angle$ DBA, he writes $\angle$DBA + $\angle$BAC + 90° = 180° and therefore $\angle$DBA = 180° - 90° - $\angle$BAC = 90° - $\angle$BAC	Evidence generation
17	Having written that down he pauses for a moment, reconsidering what he just wrote because he had the feeling something was odd.	Evidence evaluation
18	After rereading what he had just written and looking at the sketch,	Evidence evaluation
19	he crosses out both lines because he realized that he does not know for sure that $\angle$AMB = 90°.	Drawing conclusions
20	Having failed to use the angles, he looks at the sketch again, this time focusing more on "what he really knows".	Evidence generation
21	That basically amounts to $\overline{AB} = \overline{BC}$ and $\overline{AD} = \overline{DC}$.	Evidence generation
22	That in mind, he realizes that both "upper and lower" triangle have the diagonal BD in common, as well as two sides of the same length.	Evidence generation
23	He mumbles: Aha, that's how it works. It's true!	Drawing conclusions
24	He therefore writes down: The triangle ABD and BCD have three sides of the same length.	Evidence generation
25	Per definition $\overline{AB} = \overline{BC}$ and $\overline{AD} = \overline{DC}$.	Evidence generation
26	The third line BD is literally the same in both triangles, so it has the same length.	Evidence generation
27	Knowing this, we can use the "side-side-side" rule to see that both triangles are congruent.	Evidence generation
28	But then, as congruent triangles have the same angles, the angels $\angle$BAD and $\angle$DCB must be the same, because they are the corresponding angles in both triangles.	Evidence generation
29	Rereading what he has written,	Evidence evaluation
30	Luke is satisfied and concludes that his hypothesis was correct and is now proven.	Drawing conclusions

3 Current State of Research

Outline *This chapter reviews current research related to mathematical argumentation and proof skills. First, we introduce a framework to characterize mathematical argumentation and proof skills as a complex cognitive skill. The framework combines three different research aspects; the individual resources underlying the skill, the processes the skill is enacted in and the situations that require the skill. Each of the three aspects is then regarded in-depth and current research findings relevant to this project are presented.*

Mathematical argumentation and proof have always been among the most prominent topics within mathematics education and neighboring disciplines such as philosophy of mathematical practice (e.g., Hanna, 2000; Lakatos, 1976). Numerous papers have been dedicated to the importance of mathematical argumentation and proof, conceptions of proof, aims of proofs, as well as on students' shortcomings related to proofs (e.g., Boero, 1999; de Villiers, 1990; Duval, 1992; Hanna, 1990; Harel, 1999; Jahnke, 2007; Lakatos, 1976; Reid & Knipping, 2010; Stylianou, Blanton, & Rotou, 2015; Thurston, 1994; Yackel & Cobb, 1996). Yet, in her review, Mariotti (2006, p. 173) comes to the conclusion that there is

> a move away from early studies, focused on students' (and more rarely teachers') conceptions of proof, and generally speaking on difficulties that pupils face in coping with proof and proving, towards more recent studies where researchers present and discuss opinions on whether and how is it possible to overcome such difficulties through appropriate teaching interventions.

Alongside this shift in research, three other changes in research on mathematical argumentation and proof can be seen:
First, there is an increased focus, especially in empirical research, on argumentation and proof in higher education, that is in (under)graduate mathematics (e.g., Andriessen, 2009; Attridge & Inglis, 2013; Biggs, 1989; Jaworski, Treffert-Thomas, & Bartsch, 2009; Jones, 2000; Rach & Heinze, 2011; A. Selden, 2011; Trenholm, Alcock, & Robinson, 2016). Here, especially the transition from secondary to tertiary education (e.g., M. Clark & Lovric, 2008; Corriveau & Bednarz, 2017; De Guzmán, Hodgson, Robert, & Villani, 1998; Gueudet, 2008), the starting and undergraduate phase at university (e.g., Alcock, Attridge, Kenny, & Inglis, 2014; Alcock & Inglis, 2009; Maciejewski & Star, 2016; Moore, 1994; Rach & Heinze, 2013; Rach et al., 2014; Reichersdorfer et al., 2014), as well as the transition to proof-based mathematics courses are of interest.[5]
Although the focus on higher education and undergraduate mathematics is a general trend, mirrored in the *Research in Undergraduate Mathematics Education* (RUME) conference with its 20th anniversary in 2017, the foundation of journals like the *International Journal of Research in Undergraduate Mathematics Education* in 2015, or the creation of research centers like the *Kompetenzzentrums Hochschuldidaktik Mathematik* (Centre for Higher Mathematics Education) in 2010, this trend has particular influence on mathematics education research regarding mathematical proof at university.

[5] It should be noted that the transition from secondary to university education coincidences with the transition to proof-based mathematics in Germany and several other countries. Yet, for example, in the United States of America, the transition to proof-based mathematics will only occur later after taking basic calculus courses.

Second, there is a shift away from conceptual questions about proof towards the processes employed when handling mathematical argumentation and proof (see Weber, 2004). This can be seen on a conceptual level, for example by Reid and Knipping (2010) distinguishing between research on *proof* and *proving*, and on an empirical level, where an increasing number of studies analyze argumentation and proof construction process (e.g., Carrascal, 2015; Kirsten, 2017; Koichu & Leron, 2015; Ottinger, Kollar, & Ufer, 2017; Sandefur, Mason, Stylianides, & Watson, 2013; Van Spronsen, 2008; Watson, Sandefur, Mason, & Stylianides, 2013) (c.f. section 3.3).
Last, alongside this second change, the third shift is the acknowledgment that mathematical argumentation and proof skills are used in a wider variety of situations than the construction of novel proofs. Reading, comprehending, and validating proofs, as well as the (pedagogically effective) presentation of proofs, play an increasingly important role in mathematics education research (e.g., Alcock & Wilkinson, 2011; Hodds, Alcock, & Inglis, 2014; Inglis & Alcock, 2012; Lai & Weber, 2014; Lai, Weber, & Mejía-Ramos, 2012; Mejía-Ramos & Inglis, 2009a; Roy, Alcock, & Inglis, 2010; A. Selden & Selden, 2003, 2015a; Yang & Lin, 2008).

3.1 Combining Perspectives on Mathematical Argumentation and Proof Skills

A consequence of these shifts is the development of multiple research perspectives focusing on different aspects of mathematical argumentation and proof skills. One way to integrate these perspectives is to conceptualize mathematical argumentation and proof skills as a complex cognitive skill, that is as a latent cognitive (and partially affective-motivational) disposition underlying a person's performance (measured relative to given norms) in certain situations (e.g., Klieme & Leutner, 2006; Koeppen et al., 2008; van Merriënboer, Jelsma, & Paas, 1992; Weinert, 1999)[6]. Accordingly, mathematical argumentation and proof skills constitute an individual disposition, which is not directly accessible for researchers, but determines students' performance in situations that require mathematical argumentation and proof skills. In turn, as a latent construct students' mathematical argumentation and proof skills depend on the availability of several individual resources. This conceptualization has been proposed by several theoretical and empirical accounts within and outside of mathematics education (e.g., Chinnappan et al., 2012; De Corte et al., 2000; Schoenfeld, 1985; Ufer et al., 2008). The idea of underlying resources is based on the elementary insight that in order to handle a mathematical proof, a person is required to have a certain amount of mathematical content knowledge (e.g., definitions of the objects within the proof), may need to employ various problem-solving heuristics when constructing a proof, and needs to constantly monitor the progress during proof construction. A well-known example of a framework incorporating the idea of underlying resources is the framework by Shulman (1986) for teaching skills, introducing multiple distinct underlying knowledge facets including for example pedagogical content knowledge and pedagogical knowledge.

[6] The definition of *complex cognitive skills* used here is analogue to the concept of a *competence* used in European educational research. As the concept of competence is less familiar in English-speaking research, we chose to utilize the word *complex cognitive skill* for this project. Both concepts are used mainly interchangeably in this thesis, yet the concept of a competence may explicitly include affective-motivational resources, whereas this is not explicitly stated for complex cognitive skills. Still, for example Schoenfeld (1985) or De Corte, Verschaffel, and Op ' t Eynde (2000) include corresponding resources (e.g., belief systems) in their models for complex cognitive skills.

Mathematical argumentation and proof skills can therefore be regarded as dependent on a set of underlying resources, which are needed to successfully handle argumentation and proof in various situations. Thus, the success of mathematical argumentation and proof activities is dependent on the availability of these underlying resources and on the specific demands that are posed by the situation the argumentation and proof skills are needed in. Still, this set of resources itself and the relative influence of each resource of this set on overall mathematical argumentation and proof skills, respectively students' performance in a given situation, are largely unknown (see section 3.2).

The nature of mathematical argumentation and proof skills as a complex cognitive skill, its relation to underlying *resources*, as well as the *processes* and *situations* it is connected to, can be conceptualized using a recent framework by Blömeke et al. (2015) that focuses on the assessment of complex cognitive skills in higher education and is embedded in research on teaching skills. The framework combines the three perspectives, highlighting their connections, and models complex cognitive skills in a continuum of three layers (Figure 13), so that different cognitive and affective-motivational dispositions underlie certain situation-specific skills, which in turn lead to observable behavior. Here, the term „situation-specific skills" is used for *processes* connecting the dispositions underlying the complex cognitive skill and the performance (Blömeke et al., 2015, p. 7). These are internal processes that are not directly accessible for researchers or external observers, however some characteristics of these processes may be inferred from students' behavior, interactions, verbal utterances, or other observable processes.

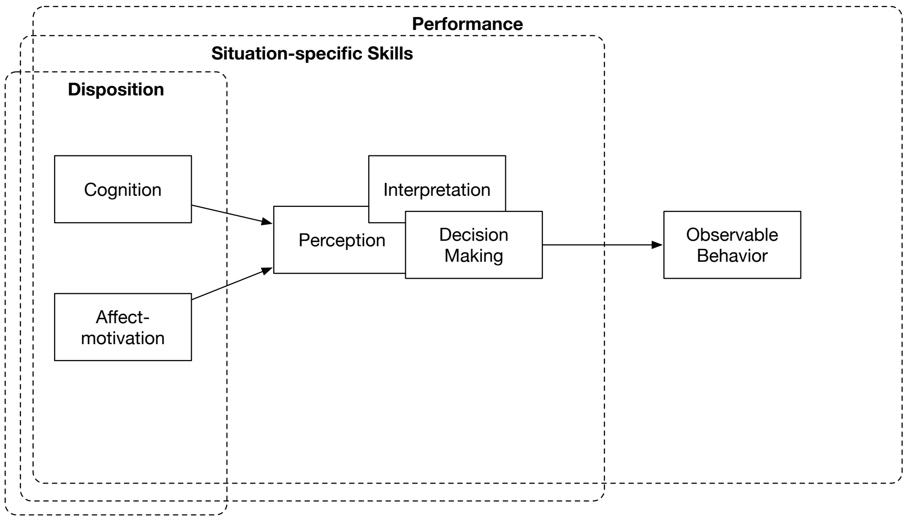


Figure 13. Framework of Blömeke et al. (2015) connecting multiple aspects of complex cognitive skills.

The framework thus connects individuals' *resources* with their observable performance in certain *situations*, which is modulated (see also Schoenfeld, 2010) by specific *processes*.

Following this perspective, a decent understanding of mathematical argumentation and proof skills has to involve knowledge about the resources and processes required in order to be successful in various situations that require mathematical argumentation and proof skills, as well as their relation among and between each other.

3.2 Resources

So far, no general framework for the resources underlying mathematical argumentation and proof skills exists, and no definite list of resources has been given. In contrast, there are several frameworks related to the resources underlying problem solving, self-regulated learning, and proof construction. These frameworks as well as several individual studies can be used to identify and characterize potential resources that are likely to underlie mathematical argumentation and proof skills. In doing so, the assumption is made that the resources underlying problem solving or self-regulated learning may also be underlying mathematical argumentation and proof skills in general. This seems likely, as handling mathematical argumentation and proof can be seen as a specific kind of problem solving and also requires self-regulation. The assumption is empirically validated in this project as the status of the potential resources is empirically checked in our correlational study (see section 5.2).

Schoenfeld (1985) introduced a framework consisting of four different resources required for successful mathematical problem solving (Figure 14): Besides various forms of mathematical knowledge such as facts and algorithmic procedures, which he terms *resources*[7], general problem-solving *heuristics* (e.g., Abel, 2003; Gigerenzer, 2008; Simon, 1978; Todd & Gigerenzer, 2001), that is domain-general strategies and techniques that can be used for unfamiliar problems, are needed. He further mentions that *control* mechanisms are needed, such as planning and monitoring, as well as *belief systems,* for example about one-self, mathematics, or the topic.

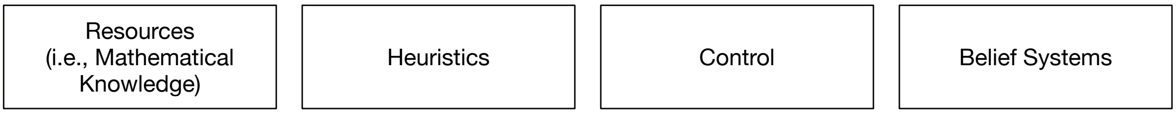


Figure 14. Resources underlying problem-solving skills according to Schoenfeld (1985).

Although labeled differently, the same resources are found in the framework by De Corte et al. (2000) in the context of self-regulated learning, which they label "categories of aptitude required to obtain a *mathematical disposition"*. Besides a domain-specific knowledge base, heuristic methods, and beliefs related to mathematical learning and problem solving, they mention *metaknowledge* as well as *self-regulatory skills,* and thereby describe Schoenfeld's *control* category more closely.

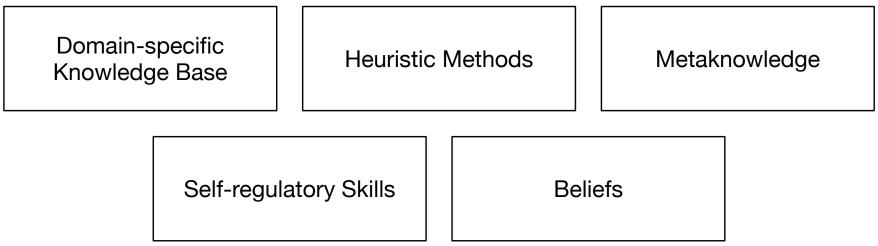


Figure 15. Resources required for a mathematical disposition according to De Corte et al. (2000).

The framework of resources for problem-solving skills by Carlson and Bloom (2005) can be regarded as a next step in the development of these frameworks, as it is based on an extensive literature review, complemented by a qualitative analysis of mathematicians problem-solving behavior to identify those resources that "occurred consistently in the problem-solving process" (Carlson & Bloom, 2005, p. 66). Here, Carlson and Bloom inferred the resources employed while solving problems from mathematicians' observable behavior and oral statements. As result of

[7] The term „resources" in Schoenfeld's framework is used to describe knowledge facets only, whereas it is used to describe any kind of disposition underlying a complex cognitive skill within this project.

their work they list *resources*, that is conceptual knowledge, facts, and algorithms, *heuristics* such as drawing a graph or computational heuristics, *affect* including beliefs and motivation, as well as the skill to *monitor* one's own behavior including self-talk or other reflective behavior (Figure 16). Later K. Clark, James, and Montelle (2014) suggested adding a fifth category related to a person's *behavior within groups*.

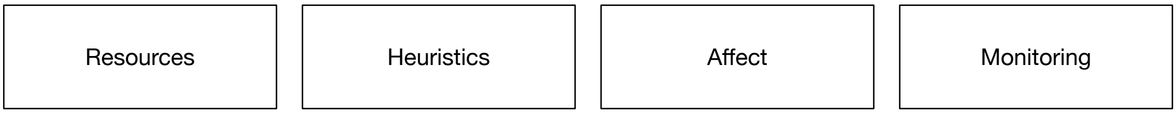


Figure 16. Resources of problem-solving skills according to Carlson and Bloom (2005).

Complementing these more general frameworks, two frameworks are drawn from quantitative studies in the context of proof construction in secondary school geometry classrooms: In their study, Ufer et al. (2008) were able to show that *declarative* and *procedural mathematical knowledge* as well as *problem-solving skills* contributed significantly to the explanation of students' performance in proof construction (Figure 17).

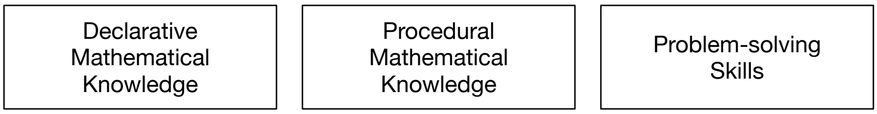


Figure 17. Resources quantitatively demonstrated by Ufer et al. (2008) in the context of secondary school geometry proof construction.

Chinnappan et al. (2012) replicated the majority of these results, again finding a significant contribution of *content knowledge* and *general problem-solving skills* (Figure 18). Furthermore, they showed an impact of *mathematical reasoning skills* conceptualized as a "broad range of reasoning skills that students activate in the context of solving a range of classroom mathematics problems" (Chinnappan et al., 2012, p. 3), which were operationalized by using students' grade 10 final scores.

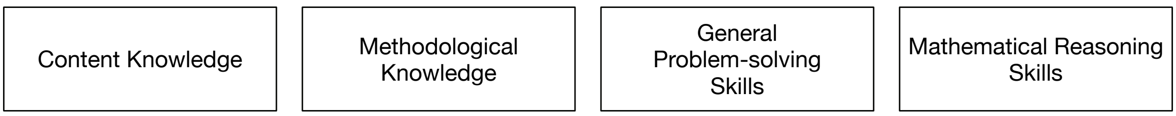


Figure 18. Resources quantitatively demonstrated by Chinnappan et al. (2012) in the context of secondary school geometry proof construction.

Not included in Chinnappan et al.'s quantitative study (2012) but emphasized throughout their paper as a fourth resource is *methodological knowledge*, which has been proposed by Heinze and Reiss (2003) based on a study of proof validation in secondary school classrooms. Following the definition of Heinze and Reiss, methodological knowledge is knowledge about various types of proofs, their nature and purposes, as well as acceptance criteria for mathematical proofs.

Despite the number of resources already presented within these relatively exhaustive frameworks, several other resources have been proposed over time. For example, Raman (2003) suggested a persons' *epistemologies*, which have not been explicitly mentioned in any of the reviewed frameworks. Furthermore, Weber (2001) suggested *mathematical strategic knowledge* as a resource, that is knowledge about cues and hints within mathematical tasks that lead to promising concepts and methods for tackling the task. Evidence for this was subsequently found by several studies (e.g., Reiss & Heinze, 2004; A. Selden & Selden, 2008). Also *conditional* *reasoning skills*, which are supposed to be essential for deductive reasoning and therefore especially for the handling of proof, have been proposed (see Alcock, Bailey, Inglis, & Docherty, 2014; Epp, 2003; Johnson-Laird, 2000; Johnson-Laird & Byrne, 2002).

On a meta-level, the comparison of the frameworks and studies highlights several points about resources underlying mathematical argumentation and proof skills:

1. The resources differ in their domain-generality: Whereas content knowledge is restricted to a specific mathematical content area, methodological knowledge is valid across mathematical content areas, and problem-solving heuristics may even be used across various domains.
2. The resources comprise various types of knowledge and skills: Examining the resources by prior research, conceptual, procedural, and strategic knowledge facets (de Jong & Ferguson-Hessler, 1996) as well as several different skills can be identified.
3. The resources depend heavily on their categorization and labeling: Each of the frameworks uses slightly different terms (i.e., mathematical knowledge base, resource, declarative and procedural mathematical knowledge) and conceptualizations, leading to a varying number of resources as well as varying conceptual breadth of the individual resources.
4. Empirical results regarding the influence of a certain resource largely depend on its conceptualization and operationalization: Although the results of both quantitative studies (Chinnappan et al., 2012; Ufer et al., 2008) point into similar directions, the differences regarding resources and conceptualizations used in the frameworks subsequently also lead to partially differing findings.
5. Although various frameworks exist, many based on a certain amount of empirical evidence, the importance of the individual resources for mathematical argumentation and proof skills is largely unknown. Here, the studies by Ufer et al. (2008) and Chinnappan et al. (2012) have given first insights, but they are limited in their scope as multiple potential resources were not included.
6. Although combining multiple frameworks and studies in this review on resources underlying mathematical argumentation and proof skills, this list of resources is most likely incomplete and incorporates only those resources that are deemed to have a substantial impact.
7. So far, most research regarding resources has been focusing on problem solving or proof construction. For other situations such as reading a proof, little information regarding the underlying resources is available, mostly originating from corollary findings or based on similar activities in other domains (Bleiler, Thompson, & Krajcevski, 2014; Chi, Bassok, Lewis, Reimann, & Glaser, 1989; Inglis & Alcock, 2012; Ko & Knuth, 2013; Mokhtari & Reichard, 2002; A. Selden & Selden, 2003; Weber & Mejía-Ramos, 2011).

3.3 Processes

For researchers, the processes involved in handling mathematical argumentation and proof are of particular importance. As mathematical argumentation and proof skills themselves are a latent construct and therefore are not accessible, the enacted processes offer a primary access to examine and evaluate the underlying skill beyond the mere consideration of performance. This approach has been used much in qualitative studies (e.g., Knipping, 2008; Koichu & Leron, 2015; Pfeiffer, 2010; J. Selden & Selden, 2009; Smith, 2006; Weber, 2004, 2008) and is increasingly receiving attention by quantitative research as modern technology facilitates to observe processes efficiently in the context of mathematical argumentation and proof skills even at a large scale (e.g., Kirsten, 2017; Ottinger et al., 2017; Ufer, Heinze, & Reiss, 2009b; Vogel et al., 2016).

Still, up to now research regarding processes in the context of mathematical argumentation and proof skills has mainly focused on proof construction and problem solving in general and only first insights into other situations such as proof comprehension exist (e.g., Alcock & Weber, 2005; Weber, 2008; Weber & Mejía-Ramos, 2011; Weber, Mejía-Ramos, Inglis, & Alcock, 2013). Accordingly, current research on processes in the context of mathematical argumentation and proof rather cover only certain aspects. Moreover, it is partially unclear which processes mediate between students' individual resources and their performance in various situations.

Three process frameworks for problem solving, proof construction, and scientific reasoning and argumentation are laid out in the following and compared regarding their individual processes as well as the larger phases within these frameworks.

3.3.1 Problem Solving

Constructing mathematical proofs is often regarded as a problem-solving activity and seminal work on problem solving had a strong focus on this activity. A classical model resulting from this research that has often been used, mainly in slightly modified forms, in the context of mathematical argumentation and proof (e.g., Kapa, 2001; Kelly, 2006; Mason, 1982; Nunokawa, 1994) is Polya's (1945) framework for problem solving. It comprises four different phases describing an idealized problem-solving process (Figure 19) and starts with the *understanding* the problem. After that, the problem solver first should *devise a reasonable plan* for tackling the problem and subsequently *carry out the plan*. Having done this, the problem solver is supposed to *look back* at the results and check their correctness. Within Polya's framework, the problem solver is allowed (and probably expected) to switch back and forth between these phases multiple times, for example when getting stuck in carrying out a plan and having to devise a new plan, or shifting from looking back to planning after noticing that the result is wrong.

Understanding the Problem	Devising a **Plan**	**Carrying out** the Plan	**Looking Back**

Figure 19. Phases while solving problems (Polya, 1945).

Although aimed mainly at mathematical problem solving, the formulation of the phases (Figure 19) denotes the general applicability of Polya's framework. Its phases can be used across many disciplines, be applied to different sorts of problems, and be used for example across all epistemic modes (see section 2.4.3). Still, the framework has to be adapted and elaborated for specific situations, especially when attempting to code individual, observable processes for research purposes.

3.3.2 Proof Construction

In contrast to Polya's general approach, the framework created by Boero (1999) focuses explicitly on the construction of mathematical proofs. It consists of six phases aiming to describe mathematics experts' behavior when constructing a mathematical proof (Figure 20). These phases can be divided into two larger types of phases: a conjecturing and exploration phase (including the *production of a conjecture*, *formulation of the statement* and *exploration of the content*) and a phase directed more closely to constructing the argumentation and final proof (including *selection and enchaining of coherent theoretical arguments into a deductive chain*, *organization of the enchained arguments into a proof* and *approaching a formal proof*).

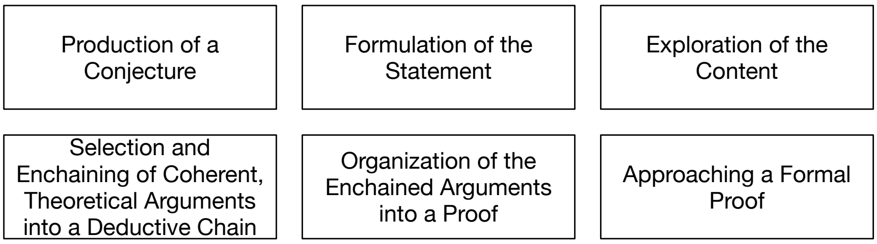


Figure 20. Phases while constructing mathematical proofs (Boero, 1999).

According to Boero, the six phases are not meant to be in linear order, nor can they be clearly separated in every case. He also highlights that the phase of approaching a formal proof may not be part of every proof construction process as a *formal* proof is often neither desired nor achievable within a reasonable amount of time and work. In comparison to the framework by Polya (1945), Boero's (1999) framework stresses the construction of a mathematical proof out of the initial argumentation by allocating three of his phases to this part. Further, Boero omits Polya's last phase of reflecting and looking back.

3.3.3 Scientific Reasoning and Argumentation

In a more general approach, the framework by Fischer, Kollar, et al. (2014) outlined in section 2.4.3 aims to describe the processes within any kind of scientific reasoning and argumentation activity. Conceptually it is therefore situated in-between the frameworks by Polya and Boero: On the one hand, it is (at least from its name) more closely focused on argumentation like Boero's framework instead of problem solving in general. On the other hand it is domain-general like Polya's framework and allows an application in a large number of settings, ranging from evidence-based reasoning of social workers (Ghanem, Kollar, Fischer, Lawson, & Pankofer, 2016) to problem solving of future teachers (Csanadi, Kollar, & Fischer, 2016).

Contrary to both Boero's and Polya's frameworks, the framework by Fischer et al. explicitly includes the process of *communicating and scrutinizing the results* and devotes a separate process to it. Fischer et al. thereby acknowledge the social context of scientific reasoning and argumentation not only in terms of social norms, but also that results of these processes are usually meant to be shared within a community where these will be subject to a social acceptance process that may lead to the rejection or refinement of the results. The first part, that is the communication, is at least partially mentioned in Boero's framework which includes "the production of a text for publication" in his fifth process (Boero, 1999, p. 3). Here, the process of communicating the results is therefore identified with writing down the proof in a readable way that corresponds to the socio-mathematical norms of the local community (e.g., including rigor definitions of each variable that may have been omitted during the construction process). Heinze and Reiss (2007, p. 342) furthermore suggested to add the seventh phase *acceptance by the mathematical community* to Boero's framework.

3.3.4 The Trichotomy of Proof Construction

Based on the common features and differences of all three frameworks a trichotomy of larger main phases of constructing a mathematical proof can be identified (Figure 21), which has also been described by Schwarz, Hershkowitz, and Prusak (2010). First, the given mathematical proposition as well as its premises have to be understood and a chain of arguments, not necessarily building on definitions and axioms, has to be created in order to *solve the problem* behind the given mathematical proposition. During this main phase, an informal argumentation is created connecting the proposition to the premise to justify that the proposition holds.

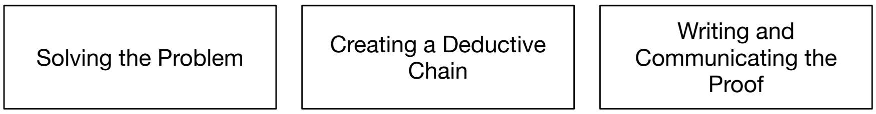


Figure 21. Trichotomy of proof construction.

Speaking in terms of Boero's framework, this refers to the phases 1 to 3, yet so far no absolutely clear line of argumentation and especially no clear deductive chain building on existing theorems is established. In Polya's terms, the problem may already be solved as Polya's framework neither explicitly includes the adaption of the solution to some standards of the community nor the communication of the solution.

A. Selden and Selden (2013) call this first phase the *problem-oriented part*, Tall (1992) the *creative phases of mathematical thinking,* and Pedemonte (2008) would argue that the *argumentative part* is over, yet the *transformation into a proof* is still missing.

This transformation of the argumentation into an (at least mental) proof represents the second main phase *creation of a deductive chain of arguments*. De Guzmán et al. (1998) call this transformation the transition from the *ordinary logic of everyday life* to the logic of mathematics. Here, the connection between the proposition and the premises gained in the first main phase are transformed into a deductive argumentative chain. In Boero's framework, this is included in the phases 4 and 5. Depending on the notion of a *formal* proof, it may also entail (parts of) phase 6, because approaching a formal proof includes filling logical gaps and reconsidering how evident the used theorems and inferences are.

The distinction between the two main phases *solving the problem* and *creating a deductive chain* mimics the views by Duval (1992) and Balacheff (1999), distinguishing clearly between argumentation and mathematical proof (French: *argumentation* and *démonstration*). It further matches the classification of proofs as being a specific kind of argumentation (see section 2.1), as the argument created in the first main phase is refined into an argumentation that complies with the local socio-mathematical norms in the second phase and thereby may be accepted as a proof, if correctly communicated in the third main phase. The distinction between the first two main phases also acknowledges the gap between argumentation and proof (Pedemonte, 2007). Despite this important distinction of both main phases, research under the term of *cognitive unity* examines how both main phases are connected and emphasizes that there is also a continuum between argumentation and proof (e.g., Boero, Garuti, & Mariotti, 1996; Pedemonte, 2007; Zazkis, Weber, & Mejía-Ramos, 2014).

Whereas the second main phase may be purely mental, rearranging and reinterpreting the solution of the problem from the first phase to obtain a deductive chain ascertaining the given claim, the third main phase, *writing and communicating the proof*, explicitly addresses the process of writing down the deductive chain achieved in the last phase according to the local socio-mathematical norms. Depending on the norms, this includes a sufficiently detailed representation of the argumentative chain, which is understandable and may entail the use of algebraic notation, correctly defined variables, definitions, and sentential connectives. This corresponds to the fifth phase in Boero's framework (*organization of the enchained arguments in a proof*), which comprises the production of a proof for publication, as well as to the eighth epistemic activity (*communicating and scrutinizing)*.

The sixth phase of Boero's framework (*approaching a formal proof*) is for the most part not included within the three main phases of the trichotomy, as formal proofs are rarely desired in mathematical practice. Boero (1999) himself mentions that "this phase may be lacking in mathematicians theorems" and refers to Thurston (1994) stating that creating completely formal proofs is "practically impossible". To include Boero's sixth phase into the model, a fourth main phase should be added.

Although the third main phase of proof construction is sometimes not emphasizes in research and practice, it may be important to examine whether students' are able to correctly write down a proof, using mathematical notation (see Ottinger, Kollar, & Ufer, 2016), in a way that corresponds to the local socio-mathematical norms. Otherwise the proof as the product of these activities may not be accepted as a valid proof by the local community, even though the underlying problem has been solved and a purely deductive argumentative chain has been (at least mentally) created (see section 2.2). For example, J. Selden and Selden (1995) showed in their study that this unpacking of informally written statements into mathematical arguments using logical symbols and variables is an obstacle for many students.

Finally, the three main phases should not be understood as a linear sequence, although this may be the case for some proof processes. Students, as well as experts, may frequently switch between all three main phases, possibly leading to synergy or discord (see further Balacheff, 1999; Boero, 1999; Boero, Garuti, & Mariotti, 1996; Pedemonte, 2007). For example, students who have advanced a bit in solving the given problem may try to write their progress down using acceptable mathematical notation and by that may be able to see what the next step in the problem-solving phase is just by looking at the written formula, text, or graph. Or the necessity to create a deductive chain may induce students to use a fixed formal scheme (e.g., complete induction), which in turn helps to solve the problem.

3.3.5 Conclusions

The frameworks, the respective processes, phases, and also main phases within the trichotomy of proof construction highlight that important processes in the context of mathematical argumentation and proof skills have been identified by prior research, that research has started to examine the processes, and that there are several current approaches to these. This is underlined by the fact that the frameworks above have been used – at least with slight adaptions to the context – successfully to describe or analyze argumentation processes, both within and outside of mathematics.

The discussion of the frameworks further shows that both Polya's (1945) and Fischer, Kollar, et al.'s (2014) frameworks are not fully capable of describing proof construction processes, as they do not explicitly incorporate the creation of a deductive proof out of the (informal) argumentation. Obviously, this is a matter of operationalization of the various processes. The second main phase of the trichotomy could be subsumed into the existing processes of the framework, but without additional sub-processes, no clear distinction between the construction of deductive and other types of arguments is possible. Thus, explorative argumentation processes, the construction of argumentations following an already existing plan, as well as the construction of a deductive chain of arguments would be described by the same processes and would, therefore, be indistinguishable, which may be desirable in some, but not all situations and research contexts. An example for this can be seen in section 2.4.3, where we showed how the epistemic activities can be used for coding the processes during the construction of mathematical argumentation and proof. In the example, the evidence generated before line 23 can be seen as part of the first main phase of the trichotomy, where the student tries to create an argument for his hypothesis. In line 23 ("He mumbles: Aha, that's how it works. It's true!"), he seems to have solved the problem and understood why his hypothesis has to hold. Thus, all evidence generated thereafter refers to the second and third main phase of the trichotomy and has to be seen as epistemologically different to the evidence created before line 23. Still, both types are coded in the same category "evidence generation".

To understand the process of mathematical proof construction in a comprehensive way, all three main phases of the trichotomy are inherently important as it does make a difference from an educational perspective whether students were not able to solve the underlying problem, to structure their solution in the form of a deductive chain, or to write down and thus communicate their proof.

Comparing the frameworks, it also becomes clear that examining the processes in the context of mathematical argumentation and proof skills can be done on different levels of precision or resolution. Whereas the framework by Fischer, Kollar, et al. (2014) can be used to describe students' processes in a very detailed way, possibly using the processes to code on the level of sentences, "units of meaning", or short time intervals (see further Strijbos, Martens, Prins, & Jochems, 2006; Strijbos & Stahl, 2007) as done in the example in section 2.4.3, the other frameworks rather correspond to broader phases, for example *understanding the problem* in Polya's framework, that involve a couple of sub-processes. Which precision of analysis is desirable, obviously is subject to the research context.

Speaking of context, the frameworks above primarily relate to the processes of an individual while solving a problem, constructing a proof or employing scientific reasoning and argumentation. Yet, this is also often done within dyads or (learning) groups. If and how the processes change in a social context, whether new processes have to be added, and which of them predict the quality of the outcome are still largely open questions (see Ottinger et al., 2017; Vogel et al., 2016).

Finally, the processes described in the frameworks further highlight a problem intrinsic to this perspective on mathematical argumentation and proof skills: The observable processes that are accessible to researchers, are only approximations of the mental processes that are the initial objects of interest.

3.4 Situations

The process frameworks laid out above are mainly concerned with, or applicable to, the construction of mathematical argumentation and proof. Yet, mathematical argumentation and proof skills go beyond these specific situations. Based on the framework by Giaquinto (2005) who suggested three general mathematical activities (*making it*, *presenting it*, and *taking it in),* Mejía-Ramos and Inglis (2009a, 2009b) structure the situations that comprise the domain in which mathematical argumentation and proof skills are applied.They describe the situations in the context of mathematical argumentation and proof in terms of three central argumentative activities (Figure 22, ovals), namely *constructing a novel argument, reading a given argument,* and *presenting an available argument,* which are also often named *proof construction, proof reading*, and *proof presentation* (e.g., Hodds et al., 2014; Roy et al., 2010; A. Selden & Selden, 2015a; Weber, 2004). Each of the three activities entails several sub-activities (Figure 22, rectangles), which Mejía-Ramos and Inglis describe in more depth according to their givens, goals, and products (see also Table 1) based on the work of de Villiers (1990).

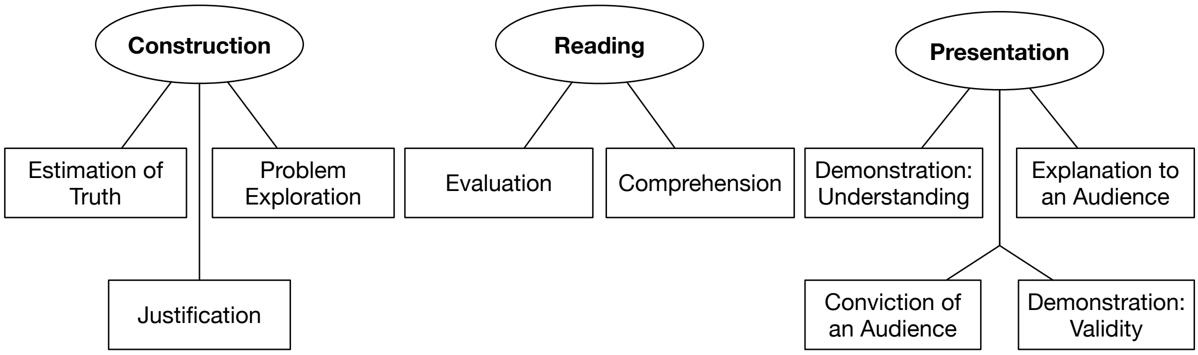


Figure 22. Framework of argumentative activities by Mejía-Ramos and Inglis (2009a).

According to their framework, *constructing a novel argument* refers to situations where either a given problem has to be *explored* in order to derive a new claim, a novel argument is constructed to *estimate the truth* of a given claim, or a *justification* for a claim in the form of a proof has to be constructed. Among these three, the term *proof construction* corresponds best to the last situation, as here a given claim, which can be assumed to be correct, has to be justified with a mathematical proof. Thus, proof construction in their framework stands for the attempt to construct a valid proof for a given claim, which corresponds to the socio-mathematical norms of the local community (A. Selden & Selden, 2015a) and thereby matches the definition used in this thesis (see section 2.1).

Table 1. Detailed description of activities related to reading arguments (Mejía-Ramos & Inglis, 2009a, p. 90).

	Reading a Given Argument	
	Comprehension	**Evaluation**
Given	An argument	An argument and a set of criteria
Goal	Understand the argument	Assess the argument against the given criteria
Product	Possibly sub-arguments with the given argument's statements as claims	An assessment (yes/no or continuous) and possibly a justification of the assessment

Within *reading a given argument*, Mejía-Ramos and Inglis (2009a, 2009b) distinguish two different activities, namely *comprehension* and *evaluation* (Table 1). Both differ in all three categories of the framework (givens, goals, and products): For example, for the activity of comprehension only an argument or proof is given, whereas for the evaluation a set of criteria is given in addition, reflecting, for example, the socio-mathematical norms (see also section 2.2). Furthermore, the goals of both activities differ between *understanding the argument* for comprehension and *assessing the argument against the given criteria* for evaluation.
Trying to align terminology in current research on mathematical proof based on a small literature review, A. Selden and Selden (2015a) add a further facet to this differentiation. They divide proof evaluation into the two activities of *proof validation* and *proof evaluation*. Both activities differ slightly in the goal they pursue: Whereas proof validation purely tries to assess the validity of the given argument as a mathematical proof, that is if it complies with the given norms, proof evaluation additionally encompasses other criteria such as beauty, elegance, or novelty. Although proof validation is included in proof evaluation in the sense that every proof evaluation also includes a proof validation, we use both terms in this thesis. First, the notion of proof

validation better matches what students usually are asked to do in school and university education, namely to validate the correctness of a proof and not to judge, for example, its beauty. Second, the “additional” criteria used for proof evaluation are often quite vague (see Inglis & Aberdein, 2015) and accordingly it is hard to ensure that these are handled appropriately. Thus, proof validation activities stand out from proof evaluation activities in the sense that they are more well-defined. Third, proof validation is closer related to proof construction than proof evaluation, because the process of validating one’s own inferences, arguments, and partial proofs is an important process during proof construction (see frameworks for processes in section 3.3), whereas determining the beauty or elegance may not be the primary criterion during proof construction.
Finally, *proof presentation* comprises four sub-activities, which mainly differ regarding their goals, which vary from convincing other, demonstrating one’s own understanding, demonstrating the validity of the given proof, to explaining the proof to the audience.
Overall, a literature review (Mejía-Ramos & Inglis, 2009a) based on the framework showed that current research literature from mathematics education focuses largely on construction (approx. 63% of the reviewed articles), to a smaller extent on reading (approx. 18 %) and none of the reviewed articles focused on presentation. Again, argument and proof construction is shown to be predominant in research. Yet, the proportion of papers (18%) on proof reading underline the already-mentioned increasing interest in proof reading (e.g., Alcock & Wilkinson, 2011; Conradie & Frith, 2000; Hodds et al., 2014; Inglis & Alcock, 2012; Lin & Yang, 2007; Mejía-Ramos, Fuller, Weber, Rhoads, & Samkoff, 2012; Mejía-Ramos & Weber, 2014; A. Selden & Selden, 2015a; Weber, 2015; Weber et al., 2014; Weber & Mejía-Ramos, 2011; Weber et al., 2013; Yang, 2012; Yang & Lin, 2008). These findings and developments further correspond to the fact that recent secondary school curricula not only contain argument and proof construction, but also their validation. The US Common Core State Standards Initiative (2010) speaks of the ability to “construct viable arguments and critique the reasoning of others” and also the German KMK standards (2012) encompass the comprehension and validation of given mathematical claims as part of mathematical argumentation. Similar inclusions of proof validation can be found increasingly worldwide.

3.4.1 Proof Construction

The construction of novel proofs is often conceived as a central learning goal (e.g., Blanton, Stylianou, & David, 2003; Hanna, 2000; Stylianou et al., 2015). Thus, mathematics education research has a particular focus on this activity (see also Mejía-Ramos & Inglis, 2009a; Mejía-Ramos & Inglis, 2009b) and many researchers have contributed to the topic. One line of research is concerned with the quantitative assessment of students’ performance in constructing novel proofs and the analysis of underlying mechanisms (e.g., Healy & Hoyles, 2000; Heinze, Reiss, & Rudolph, 2005; A. Selden & Selden, 2012; Senk, 1985; The International Commission on Mathematical Instruction (ICMI), 1966; Weber, 2001, 2003) showing that students from secondary school as well as from university have difficulties with mathematical proofs in general and proof construction in particular. These difficulties with handling mathematical proofs are often regarded as one important cause for high drop-out rates (e.g., Dieter & Törner, 2012; Heublein, Richter, Schmelzer, & Sommer, 2012; Heublein et al., 2014; Seymour & Ferrare, 2015; Seymour & Hewitt, 1997) in mathematics degree programs at the university level (Jones, 2000), especially within the first semesters.
One reason for students’ difficulties that is put forward frequently is the so-called abstract character of proof-based mathematics (see section 2.1). For example, Mamona-Downs and Downs (2005) point out that students are confused by mathematical proofs and that this

confusion constrains students in constructing and handling mathematical proofs. Dieter and Törner (2012) speak of an *abstraction shock* and others mention a *distance to real life* and the shift to other representational systems (see Goldin, 1998). Based for example on didactical proof concepts, proof per se does not require abstraction[8] as for example created by mathematical notation. Still, the desired degree of generality, which is often achieved using algebraic mathematical notation, and other constraints by the local socio-mathematical norms, often lead to this perceived abstraction. There is not a single cause for this abstract character of mathematical proofs but multiple (e.g., M. Clark & Lovric, 2008; Dörfler & McLone, 1986; Healy & Hoyles, 2000; Mariotti, 2006; Moore, 1994; Rach & Heinze, 2016; Weber, 2012; Winter, 1983; Zaslavsky, Nickerson, Stylianides, Kidron, & Winicki-Landman, 2012):

1. The mathematical, formal-symbolic notation and representation used.
2. The sole reliance on deductive arguments based on an axiomatic theory.
3. The epistemological discrepancy between proofs students are asked to construct and students' conceptions of proof, as they often lack a "need for proof" and are unaware of the various goals behind mathematical proof.

A second line of research is devoted to comparisons between the behavior of novices and experts when handling mathematical proofs. In particular, it is examined how mathematical proof is taught, how instruction can help to align the behavior of novices with that of experts, and if this alignment is beneficial for students' performance (see Weber, 2009). For example, Schoenfeld (1992) compared the activities of students and a mathematics faculty member during working on a non-standard mathematics problem showing that the expert spend a substantial amount of time for making sense of the problem whereas the students almost immediately chose an approach to tackle the tasks and kept working on it although not making progress. Further, Weber (2001) compared undergraduate students' to doctoral students' proof construction processes, concluding that the primary cause for undergraduate students problems may be a lack of mathematical strategic knowledge.

Partially based on this last finding of Weber (2001), a third line of research regarding proof construction examines the proof production styles or strategies used when constructing proofs: Based on research on representation systems by Goldin (1998), Weber and Alcock (2004, 2009) introduced the distinction between *syntactic* and *semantic*[9] *reasoning* or *proof production* styles. A syntactic proof production is characterized by the predominant use of the "*representation system of proof"* (see Weber & Alcock, 2009), including, for example, formal-symbolic notation, definitions, and first-order-logic. Using this proof production style, proving is based on manipulating definitions, equations, or other relevant data (Weber & Alcock, 2004). In contrast, a semantic proof production style also regards representations of mathematical concepts from other representation systems, in particular including less formal descriptions and interpretations such as examples, graphs, sketches, or gestures that are used as a basis for producing the proof (Weber & Alcock, 2004). Figure 23 and Figure 24 illustrate both proof production styles for two real analysis tasks.

[8] Examples that proof as a concept does not rely on mathematical notation are illustrated by Byrne (2010) or Nelsen (1993). For a discussion about the nature of formalism for mathematical proof see for example (Tanswell, 2012).

[9] The semantic proof production style was later also termed referential proof production (e.g., Alcock & Weber, 2010) to highlight the use of reference objects such as examples while constructing proofs (see further Lockwood, Ellis, & Lynch, 2016).

Prove that the sum of two real convergent sequences is convergent

Proof:

Let $(a_n)_{n\in\mathbb{N}}$ be a real convergent sequence.

Therefore $\exists a\in\mathbb{R}: \forall \varepsilon'>0\ \exists N_1\in\mathbb{N}\ \forall n>N_1: |a_n-a|<\varepsilon'$

Let $(b_n)_{n\in\mathbb{N}}$ be a real convergent sequence.

Therefore $\exists b\in\mathbb{R}: \forall \varepsilon''>0\ \exists N_2\in\mathbb{N}\ \forall n>N_2: |b_n-b|<\varepsilon''$

Let $\varepsilon>0$.

$$|(a_n+b_n)-(a+b)| = |(a_n-a)+(b_n-b)| \leq |a_n-a|+|b_n-b| =: (*)$$

$$\left.\begin{array}{l}\exists N_1\in\mathbb{N}: |a_n-a|<\frac{\varepsilon}{2}\ \forall n>N_1\\ \exists N_2\in\mathbb{N}: |b_n-b|<\frac{\varepsilon}{2}\ \forall n>N_2\end{array}\right\}\text{Let } n>\max\{N_1,N_2\}$$

$$\Rightarrow (*) < \frac{\varepsilon}{2}+\frac{\varepsilon}{2}=\varepsilon$$

□

Figure 23. Example of a proof constructed using a syntactic proof production style.

Prove, that all real convergent sequences are bounded

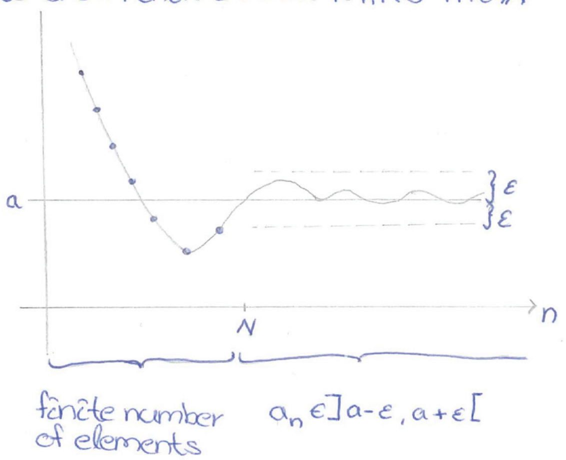
Proof:

Let $(a_n)_{n\in\mathbb{N}}$ be a real convergent sequence.

Therefore $\exists a\in\mathbb{R}\ \forall\varepsilon>0\ \exists N\in\mathbb{N}: \forall n>N: |a_n-a|<\varepsilon$

We need to show, that $\exists S\in\mathbb{R}: |a_n|<S\ \forall n\in\mathbb{N}$

Sketch:



$$\Rightarrow \text{Let } \varepsilon=1 \Rightarrow \exists N\in\mathbb{N}: |a_n-a|<1\ \forall n>N \Rightarrow |a_n| < \underbrace{|a|+1}_{\text{Boundary for "tail"}}\ \forall n>N$$

$$\text{Define } S:=\{|a|+1\}\vee\underbrace{\{|a_1|,\dots,|a_N|\}}_{\text{bound for 1st part}}$$

$$\Rightarrow |a_n|\leq S\ \forall n\in\mathbb{N}$$

□

Figure 24. Example of a proof constructed using a semantic proof production style by drawing an informal sketch of a convergent sequence.

Based on qualitative data from several prior studies, Weber and Alcock (2009) come to the conclusion that although there are many arguments for or against either proof production style and many mathematics teachers and lecturers may prefer semantic proof production, "neither strategy seems to be superior to the other strategy" for producing valid mathematical proofs. Yet, the different proof production styles may lead to different levels of conviction and understanding (Weber & Alcock, 2004). They suggest that both proof production styles can lead to robust learning about mathematical content and proof construction. Additionally, several studies (Alcock & Weber, 2010; Pinto & Tall, 1999) underline that different students favor different proof production styles and are quite consistent in doing so, at least within the same mathematical content area (Mejía-Ramos, Weber, & Fuller, 2015).
Initially, Weber (2004) proposed a third proof production styles for proof construction, which was later abandoned. This third category was the *procedural proof production*, characterized by students trying to apply a certain procedure, that is a series of steps that the student believes to be a reasonable approach for the proof. Common procedures are algorithms (e.g., using the quadratic formula to solve a quadratic equation) or specific approaches to proofs (e.g., using complete induction).

3.4.2 Proof Reading

Compared to proof construction, proof reading and the associated sub-activities have only recently attracted attention in research (e.g., Alcock & Weber, 2005; Beitlich et al., 2014; Inglis & Alcock, 2012; Lin & Yang, 2007; Mejía-Ramos et al., 2012; Weber, 2008; Yang & Lin, 2008).
One line of research regarding proof reading tries to capture students' and experts' proof reading behavior by recording and analyzing their eye-movement and gazing behavior. Although this line of research was already initiated in the 1980s (e.g., Just & Carpenter, 1980; Levie & Lentz, 1982), the broader availability of *eye-trackers* has yielded many new findings. For example, Beitlich et al. (2014) examined whether and how mathematic majors and university mathematicians use (decorative) pictorial information included in proofs during proof comprehension. Findings emphasize that participants switched back and forth between picture and text several times, likely attempting to connect and integrate both. Inglis and Alcock (2012) compared the eye-movement of beginning undergraduate students and experienced mathematicians during proof validation and found that undergraduate students spend significantly more time on surface features of mathematical proofs, a result also underlined by A. Selden and Selden (2003). Furthermore, Inglis and Alcock (2012) found that expert mathematicians shifted their attention significantly more often back and forth between different segments or lines of proofs, again possibly trying to integrate both parts and to see connections between both. Weber and Mejía-Ramos (2011) further describe two different activities when reading proof: A *zooming in* when focusing on small parts such as individual inferences or short steps of the proof, and a *zooming out* when for example looking at the overall structure of the proof. Furthermore, an initial *proof skimming* is under debate (Inglis & Alcock, 2012; Weber et al., 2013).
Building on this research on proof reading, another line of research focuses more directly on students' and experts' proof validation behavior and the ways they obtain conviction that proofs are valid (e.g., Alcock & Weber, 2005; Healy & Hoyles, 2000; Heinze & Reiss, 2003; Inglis & Alcock, 2012; A. Selden & Selden, 2003; Ufer, Heinze, Kuntze, & Rudolph-Albert, 2009; Weber, 2008; Weber et al., 2014; Weber et al., 2013). Results reveal that in spite of the conceived view, experts use authoritarian as well as other, non-deductive evidence to gain conviction (e.g., Geist, Löwe, & Van Kerkhove, 2010; Heinze, 2010; Weber et al., 2014), and that there is a mismatch between their behavior as (research) mathematicians and their behavior in educational settings, where

they often strictly disapprove the use of non-deductive evidence to gain conviction (Weber et al., 2014).

For school and university students, results show struggles with proof validation, often related to students' inappropriate proof schemes or proof conceptions (e.g., Harel & Sowder, 1998; Healy & Hoyles, 2000; Heinze & Reiss, 2003). Further, the acceptance of different types of proofs relating to educational proof concepts, for example generic proofs or visual proofs, and their educational use are currently studied (Alcock, 2009; Biehler & Kempen, 2016; Inglis & Mejía-Ramos, 2009; Kempen, 2016; Malek & Movshovitz-Hadar, 2011; Rowland, 2001, 2013).

3.4.3 Proof Presentation

Research focusing explicitly on proof presentation is scarce. So far, it has been examining mainly the presentation of proofs by teachers and lecturers, taking an educational perspective on the situation. Overall, teaching of mathematical argumentation and proof at the university level is often conceived as traditional. Many researchers (e.g., Davis & Hersh, 1981; Siebert, Rach, & Heinze, 2013; Thurston, 1994; Weber, 2012) see typical lectures in advanced mathematics organized around a *definition, theorem, and proof* structure. Mathematical proofs are then often presented not as a process, but as a product (Alibert & Thomas, 1991; M. Clark & Lovric, 2008; Siebert et al., 2013). That is, the final proof is introduced to students in a linear style, without highlighting the process of obtaining the proof, for example by illustrating the proving process or including heuristic information alongside the presentation. Thus, students are confronted with an unrealistic image of problem-solving or proof construction processes, possibly confusing them and inhibiting their proof construction (Alibert & Thomas, 1991, p. 215).

A main line of research regarding proof presentation focuses on lecturers and teachers intentions and techniques when presenting proofs (e.g., Hemmi, 2010; Weber, 2012) as well as their perspectives on what constitutes good *pedagogical proofs* (see Lai & Weber, 2014; Lai et al., 2012), that is proofs arranged and structured in such a way that they allow students without prior understanding to comprehend it. Based on an interview study, Weber (2012) highlights that most lecturers actually expressed the desire to use proofs for illustrative purposes and in an explanatory way, yet lack the pedagogical strategies and "provide little guidance to students on how to engage in the complicated process of reading and comprehension of proofs" (Weber, 2012, p. 478). Similar, in an exploratory interview study, Alcock (2010) identified four modes of thinking (*instantiation*, *structural thinking*, *creative thinking,* and *critical thinking*) that mathematics experts indicate as important for students in order to produce mathematical proofs effectively. Yet, based on the same interviews, Alcock (2010, p. 78) concludes that these experts mainly emphasize structural thinking, that is generating "a proof for a statement by using its formal structure", in courses meant to introduce students to argumentation and proof, partially distorting students' view of mathematical proof. Based on an observational study enriched by interviews, Lew, Fukawa-Connelly, Mejía-Ramos, and Weber (2016) point out another problem: Even if lecturers have best intentions and verbally give methodological and conceptual ideas as a supplement to the written down proof, students may still not take those up due to their practice of focusing on the written parts as these are supposed to be more important. Still, based on a case-study Fukawa-Connelly (2012) highlights that instructors have more educational strategies at hand than often expected.

Currently, there are several ideas how to overcome the mismatch between taught behavior and actual behavior. For example, Fukawa-Connelly (2012) advocates *structured proofs* for university lectures. Similarly, many researchers (e.g., Hilbert, Renkl, Kessler, & Reiss, 2008; Kollar et al., 2014; Reichersdorfer, 2013; Reichersdorfer et al., 2012; Reiss, Heinze, & Kessler, 2007; Reiss et al., 2008; Renkl, Hilbert, & Schworm, 2009) are working on *heuristic worked-out examples*, which mimic the proving process including steps that are not visible in the written-down proof and illustrate the use of heuristics during the proving process. Although this research so far is primarily concerned with secondary mathematics and showed mixed effects at university (e.g., Kollar et al., 2014), other ways of representing proofs and scaffolding students' engagement with proofs such as e-proofs (Alcock, 2009; Alcock & Wilkinson, 2011) or self-explanation trainings (Hodds et al., 2014) are also developed for university contexts.

3.4.4 Relations Between Situations

So far, few studies examine the interplay between proof validation and proof construction (e.g., Pfeiffer, 2009a, 2011; Ufer, Heinze, Kuntze, et al., 2009). Some studies conceptualize proof validation and proof construction as two separate skills or as two subskills of one overall mathematical argumentation skill, finding a significant, yet weak correlation between both (Ufer, Heinze, Kuntze, et al., 2009). Other research highlights that the activity of proof construction entails proof validation as a process (e.g., A. Selden & Selden, 2003), that both are intrinsically linked (Cilli-Turner, 2013), and that both can lead to reciprocal learning effects, that is "the ability to validate proofs can improve the ability to construct proofs" and vice versa (Pfeiffer, 2009a, 2011). In contrast, considerable research does not distinguish between these activities or subskills and rather speaks of proof skills in general (e.g., Hanna, 2000). As already pointed out by A. Selden and Selden (2015a), there is currently little research on the relation between the concepts of proof construction, proof reading, and proof presentation and their conceptual status remains somewhat unclear.

3.5 Educational Consequences of Conceptualizing Mathematical Argumentation and Proof Skills as a Complex Cognitive Skill

The review of current literature summarized above reveals that mathematical argumentation and proof skills can reasonably be conceptualized as a complex cognitive skill, including several resources, processes, and situations. From a research perspective, this analytic conceptualization (see Blömeke et al., 2015) is a promising approach to better understand students' mathematical argumentation and proof skills and underlying mechanisms. Moreover, according results can be – and in parts have already been – used to enrich teaching, learning, and support of mathematical argumentation and proof skills. Solid knowledge about the various aspects of students' mathematical argumentation and proof skills and their interplay may provide a sound basis to create effective means to support students. Koedinger (1998, p. 319) points out more generally:

> Developing a model of these skills is a key step toward creating effective learning environments […]. This model can then provide design guidance in creating elements of a learning environment

Thus, one consequence of acknowledging the dependence of mathematical argumentation and proof skills on several underlying resources is the fact that no longer a *solitary* skill has to be acquired or supported in learning environments, but rather *multiple* resources, as well as their combination in form of mathematical argumentation and proof skills. Based on knowledge about the underlying resources and their relative influence, those with a high impact can be explicitly supported, looking for learning gains of the resources and overall mathematical argumentation and proof skills. As for their high influence on overall mathematical argumentation and proof skills, these underlying resources represent generic candidates for first educational intervention attempts to support students. Yet, a high relative influence does not imply that these resources can effectively be supported. Here, further research is needed to examine whether these resources can be effectively supported and if, and under which condition, gains within the resources also transfer to overall mathematical argumentation and proof skills.

This *resource-based approach* to support students is especially interesting as up to now, despite content knowledge, most resources outlined in section 3.2 are (in our experience) not explicitly taught, neither at school nor university and prior research (e.g., Heinze, 2007; Mevarech & Fridkin, 2006; Perels, Gürtler, & Schmitz, 2005; Schoenfeld, 1982) suggests that students' individual resources can be effectively supported. However, so far studies mainly focused solely on one resource, for example on meta-cognition or problem-solving skills, not aiming to support multiple resources, and often only assessed learning gains regarding this one, specific resource. Thus, it is unclear whether possible effects of explicit teaching of multiple resources are restricted to resources themselves or if they transfer to overall mathematical argumentation and proof skills. Furthermore, it is unclear what instructional design should be used to organize support for multiple resources within a single course.

First hints can be found in research from instructional design that has focused on complex cognitive skills for a longer period of time. In particular, researchers have debated about the effectiveness of *part-task* and *whole-task approaches* (e.g., J. R. Anderson, Reder, & Simon, 1996; Branch & Merrill, 2011; Lim, Reiser, & Olina, 2009), trying to determine whether it is more effective to decompose a larger task into several part-tasks and teach these individually, or to directly focus on the whole task.

Based on classical learning theories (e.g., J. R. Anderson, 1996, 2002) some instructional approaches assume that complex tasks can be decomposed into part-tasks and recommend to train each of these separately. *Part-task approaches* are guided by the idea that instruction on part-tasks is of higher instructional clarity for the students, that each part-task is easier to master for learners, and that learning gains on the part-tasks easily transfer to learning gains on the overall task. Sociocultural and situated conceptions of learning (e.g., Brown, Collins, & Duguid, 1989; Greeno, 1998; Lave & Wenger, 1991; The Cognition and Technology Group At Vanderbilt, 1990) contest these approaches and highlight the situatedness of learning. Based on these conceptions of learning, *whole task-approaches* (e.g., van Merriënboer & Kester, 2007; van Merriënboer & Kirschner, 2007) reject the atomization of tasks, give evidence for the situatedness of learning, and point to difficulties associated with attempts to transfer from part-tasks to the overall task (e.g., J. R. Anderson et al., 1996; R. E. Clark & Estes, 1999; van Merriënboer, de Croock, & Jelsma, 1997).

Although the question which of both approaches is superior in which conditions is not answered definitely, the current understanding is that complex cognitive skills benefit from whole-task approaches. Transferring these results and approaches from instructional design back to mathematical argumentation and proof skills, it is reasonable to assume that teaching the various resources in a simultaneous, integrated way may be positive for overall mathematical argumentation and proof skills. Yet, it is unclear whether findings regarding part-tasks and whole-tasks can be transferred to the level of resources underlying a complex cognitive skill, as the resources have to be used non-sequentially at different points during mathematical argumentation and proof skills and thereby exhibit a higher complexity than the part-tasks mostly focused in instructional design.

Based on the approach to support the resources underlying mathematical argumentation and proof skills, foremost the support of domain-general resources is desired by some educators (e.g., D. W. Eccles & Feltovich, 2008) hoping for positive effects not only for mathematical argumentation and proof skills, but also for other skills relying on the same resource. Still, the effectiveness of domain-general trainings is debatable (see further Chinnappan & Lawson, 1996;

D. W. Eccles & Feltovich, 2008; Sweller, 1990; Tricot & Sweller, 2014) and is largely dependent on the possibility to transfer general resources acquired in one task or domain to the use in other tasks or domains (see J. R. Anderson et al., 1996; J. R. Anderson, Reder, & Simon, 1997; Greeno, 1997 for a critical discussion). It is therefore especially interesting to examine the influence of domain-general resources on mathematical argumentation and proof skills and evaluate how effective these can be supported.

Based on the suggestion by Richey and Nokes-Malach (2014, pp. 209-210) the resources underlying students' mathematical argumentation and proof skills may also be used for formative assessment. For example, at the beginning of students' mathematics studies at university a systematic analysis of the availability of these resources can inform students about their individual deficits and advantages. Also lecturers may benefit from results about the availability of resources within their classes to adapt their curricula based on this information before engaging in teaching mathematical argumentation and proof. Similarly, Schoenfeld (2012b) claims that one aim of teachers is to ensure that all students have the appropriate resources available.

Besides the resources, also acknowledging the processes may have manifold consequences for teaching. Based on future research results on the importance and impact of the individual processes that constitute performance in different situations that require mathematical argumentation and proof skills (see Ottinger et al., 2017), frameworks for the processes may allow teachers to systematically focus and cover these and thereby help students structure their individual activities. The processes can be included in instruction either explicitly by informing students about the processes, their connections, and how to regulate them, or implicitly by modeling these processes during proof construction or validation as done in analogue research on problem solving (see Heinze, 2007; Schoenfeld, 1992). Prior research has shown that the processes can also be integrated into heuristic worked-out examples (e.g., Beitlich, 2015; Hilbert et al., 2008; Kollar et al., 2014; Reiss et al., 2008). A similar approach leading to substantial effects examined by Hodds et al. (2014) is a self-explanation training that makes explicit use of the process models of proof comprehension to support students in structuring their own proof reading. Besides the benefits that students may acquire, also teachers may profit from explicit knowledge about the processes in different situations as this may have a positive impact on their professional vision (e.g., Goodwin, 1994; Sherin & van Es, 2009; Stahnke, Schueler, & Roesken-Winter, 2016), diagnostic skills (see Csapó & Szendrei, 2011; Südkamp & Praetorius, 2017), and overall teaching skills.

4 Research Framework and Guiding Questions

Outline *The first part of this chapter describes the research framework for mathematical argumentation and proof skills created for this project. We further give conceptualizations for each potential resource of mathematical argumentation and proof skills that is included in this project. Finally, we formulate and discuss the research questions guiding this project.*

Based on the work by Blömeke et al. (2015) and a growing amount of research from mathematics education (see chapter 3), mathematical argumentation and proof skills have to be seen as a complex cognitive skill entailing three important aspects: The underlying individual *resources*, the *situations* that require the use of mathematical argumentation and proof skills, and the *processes* leading to an observable performance. Each of these aspects has already received some attention by mathematics education research as well as neighboring disciplines. However, although several researchers underline that knowledge about the interplay within and between the aspects of a complex cognitive skill – such as mathematical argumentation and proof skills – is central to understand it (e.g., Blömeke et al., 2015; Heinze & Reiss, 2009; Schoenfeld, 1985, 2010; Ufer et al., 2008), research connecting these aspects is scarce.

To gain insights into mathematical argumentation and proof skills understood as a complex cognitive skill, especially focusing on the underlying cognitive resources and their relative impact, we first created a research framework including resources, processes, and situations in the context of mathematical argumentation and proof skills. Based on this framework, we conducted three consecutive studies (Figure 25):

- A literature review (see section 5.1) analyzing the aspects of mathematical argumentation and proof skills addressed by current research,
- a correlational study (see section 5.2) using Generalized Linear Mixed Models to empirically evaluate the relative impact of various resources on students' performance in proof construction and proof validation, and
- an intervention study (see section 5.3) to examine and contrast two ways to support students' mathematical argumentation and proof skills and the underlying resources.

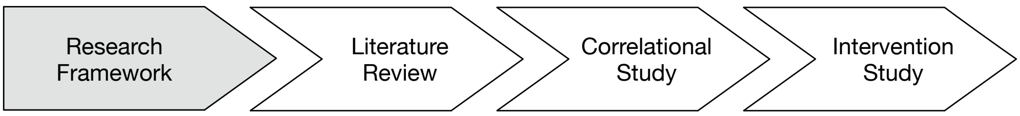


Figure 25. Illustration of the research framework and the three subsequent studies within our MIMAPS project.

4.1 Research Framework

Our MIMAPS project is based on a framework that was newly created, explicitly acknowledging the three aspects of mathematical argumentation and proof skills outlined in chapter 3. The framework is based on the work of Blömeke et al. (2015) and was adapted for the specific case of mathematical argumentation and proof skills (Figure 26) by incorporating key aspects of prior frameworks from mathematics education research (see chapter 3).

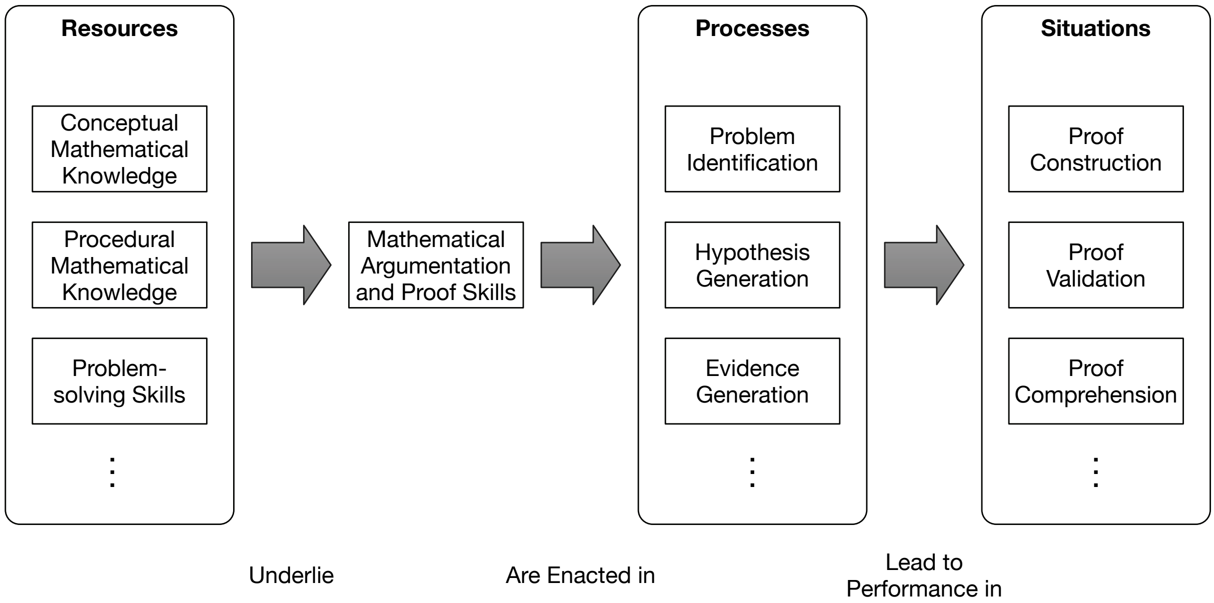


Figure 26. Research framework of our project.

In particular, we used the framework by Mejía-Ramos and Inglis (2009a, 2009b) to describe the different *situations* relevant for mathematical argumentation and proof skills[10] and the epistemic activities framework (Fischer, Kollar, et al., 2014) for the *processes* in the context of mathematical argumentation and proof. The latter framework was chosen for three reasons: First, the framework by Fischer, Kollar, et al. (2014) comprises more processes than the other frameworks allowing a finer analysis regarding each of these processes. Second, compared to the frameworks by Polya (1945) and Boero (1999) it is neither too focused on proof construction (which only is one situation in the context of mathematical argumentation and proof skills) nor too general making it hard to examine more detailed, argumentation-specific processes. Third, the framework allows connecting to domain-general research in the context of argumentation from other disciplines, facilitates meta-analyses, and is a further step towards a shared terminology. However, the selected epistemic activities framework appears to be suitable to describe only some of the situations included in our framework, namely proof construction and to a certain extent proof reading but not proof presentation. Thus, additional processes may have to be added for proof presentation at some point in future. For the purpose of our project, this limitation is acceptable, as we are not concerned with the processes of proof presentation.

The selection of a framework for students' individual resources underlying their mathematical argumentation and proof skills was more difficult, as prior research results, terminology, conceptualization, and operationalizations differed, and neither of the portrayed frameworks alone was satisfactory.

[10] Here, we used the terms *proof construction*, *proof validation*, etc. instead of the initial formulations (e.g., *argument construction*, *argument validation*), as these are more frequently used in current research, also by the initial investigators (e.g., Inglis, Mejía-Ramos, Weber, & Alcock, 2013; Mejía-Ramos et al., 2012; Mejía-Ramos & Weber, 2014; Mejía-Ramos et al., 2015; A. Selden & Selden, 2015a, 2015b).

4.1.1 Resources of Mathematical Argumentation and Proof Skills – Revisited

Based on the individual resources suggested by prior research findings (see section 3.2), a total of *nine* potential resources were selected for inclusion in our project (Figure 27), as they are distinct resources, either supported by some empirical evidence or by a strong theoretical basis. However, the included resources do not represent an exhaustive selection and the framework regarding the resources should, therefore, be rather seen as a selection of central, potential resources based on prior research findings.

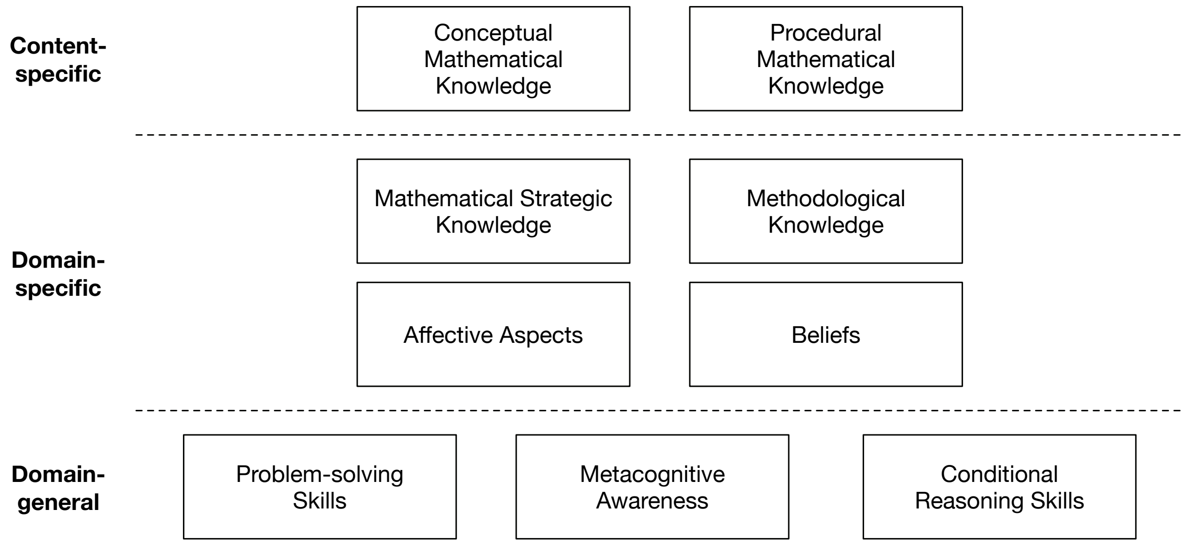


Figure 27. Potential resources of mathematical argumentation and proof skills included in this thesis.

As prior research used several, partially overlapping definitions for these resources, each will be shortly conceptualized in the following. Furthermore, we will categorize the resources regarding two research frameworks:

- Different types of knowledge included as resources are structured using the framework by de Jong and Ferguson-Hessler (1996) that resulted from a literature review. Among other things, de Jong and Ferguson-Hessler (1996) distinguish *conceptual*, *procedural*, and *strategic* knowledge. The first two are often included in different cognitive models and frameworks for knowledge (e.g., J. R. Anderson, 1982, 1983; Schneider, 2006; Star, 2005; Star & Stylianides, 2013). In contrast, strategic knowledge, sometimes also termed conditional knowledge, is not always explicitly mentioned in frameworks and, for example, sometimes included into procedural knowledge (e.g., Alexander & Judy, 1988; Baroody, Feil, & Johnson, 2007). Yet, research on problem solving (e.g, Bransford, Sherwood, Vye, & Rieser, 1986; Gok, 2010; Polya, 1945; Schoenfeld, 1985), self-regulation (Schraw, Crippen, & Hartley, 2006), and, for example, also medical education (e.g., Mayer, 2010) highlights the need for domain-specific as well as for domain-general strategic knowledge, so that we chose to include it as an individual type of knowledge as proposed by de Jong and Ferguson-Hessler (1996).
- Domain-specificity and -generality of the resources will be classified by a slightly adapted framework by Chinnappan and Lawson (1996). Within their framework, Chinnappan and Lawson (1996) distinguish between *task-specific*, *domain-specific,* and *domain-related* resources. However, as task-specific appears too narrow for our research purpose, the resources are categorized as *content-specific*, that is belonging to a specific mathematical

content area such as infinite series, *domain-specific*, that is belonging to the field of mathematics, and *domain-general* resources, that is being not specific to mathematics and hence being applicable also outside of mathematics.[11] Obviously, content-specific resources are also domain-specific, but at least when conceptualizing the different included resources we wanted to highlight this difference, which will not play a role in later analysis of data.

In addition, we will provide an example for each of the resources within a university mathematics setting to illustrate how they will be operationalized within this project (see also the individual descriptions within the studies (chapter 5)).

4.1.1.1 Resource 1: Conceptual Mathematical Knowledge

Conceptual mathematical knowledge is among the resources mentioned most often in prior research. It is usually defined as knowledge about concepts, definitions, principles, propositions, and other mathematical information (see Hiebert & Lefevre, 1986, pp. 3-4; Star & Stylianides, 2013). It is knowledge consisting of individual pieces, sometimes called factual knowledge (see also L. W. Anderson, Krathwohl, & Bloom, 2001; Krathwohl, 2002), which are (more or less) closely linked to each other (see de Jong & Ferguson-Hessler, 1996). These links between the individually pieces can therefore be seen as at least as important as the individual pieces. Conceptual knowledge is often conceived as the basis for a profound understanding of topics (e.g., Schneider, 2006) as well as for mathematical argumentation and proof skills (Hilbert et al., 2008).

Conceptual knowledge has to be categorized as content-specific, yet knowledge of certain basic mathematical concepts (e.g., conceptual knowledge of arithmetic) may be perceived as relevant across different content areas or even across domains. Within this study, the focus will be on the content-specific part of conceptual knowledge, which was also used for the operationalization and measurement within this project. An example of content-specific conceptual knowledge is the definition of a Cauchy sequence, and its links to the definition of convergence and the theorem that every Cauchy sequence converges within the real numbers.

4.1.1.2 Resource 2: Procedural Mathematical Knowledge

Besides conceptual knowledge, procedural knowledge is included in this study as a second resource. It refers to knowledge about rules, procedures, action-sequences, and algorithms for solving mathematical problems (Hiebert & Lefevre, 1986, pp. 7-8). It entails content-specific aspects as well as domain-specific aspects, yet in this project the focus will be on the former. Examples are the application of the geometric series formula while rewriting an equation, or testing a convergence criterion on a given sequence. Both examples demonstrate that procedural knowledge can also entail factual knowledge (e.g., knowledge about convergence criteria). Therefore, it is sometimes hard to distinguish between conceptual/factual knowledge and procedural knowledge. Anyhow, both types are supposed to be highly interlinked (Rittle-

[11] We chose this definition of *domain-general* to account for research questioning the claim of domain-generality and advocating the term *cross-domain* (see Fischer, Wecker, et al., 2014), as in this context it is more important that the resource is not specific to mathematics, rather than making a claim about *how* general it is. This notion is also in accordance to the notion used by Chinnappan and Lawson (1996). Further, domain-general is interpreted in the sense that an according resource may be useful or applicable also outside of mathematics. If it proves to be predictive for a certain performance in certain other domains is another, empirical question.

Johnson, Schneider, & Star, 2015; Rittle-Johnson, Siegler, & Alibali, 2001) and many learning theories suggest that conceptual knowledge can be transformed into procedural knowledge (see VanLehn, 1996). Thus, the categorization of knowledge requires background knowledge about the general setting and prior experiences of the participants. Nevertheless, it is often seen to be worthwhile on a practical and conceptual level to distinguish both types of knowledge (see Rittle-Johnson & Schneider, 2014).

4.1.1.3 Resource 3: Mathematical Strategic Knowledge

Following the first two resources, mathematical strategic knowledge seems to be the direct consequence according to the framework of de Jong and Ferguson-Hessler (1996). Still, it has been scarcely included in mathematics education research and has been suggested first by Weber (2001) as a resource on its own. Weber (2001, p. 101) describes mathematical strategic knowledge as "knowledge of how to choose which facts and theorems to apply". The term "strategic knowledge" (sometimes also "conditional knowledge") is used in a variety of ways and contexts, some not addressing its domain-generality or -specificity (de Jong & Ferguson-Hessler, 1996; Schraw et al., 2006, p. 114), others using it in the context of problem solving thereby implying a domain-generality (e.g., Gok, 2010, p. 114). Here, we use the term following Weber's definition, thus corresponding to knowledge about hints and cues in mathematical task descriptions (and in later phases of proof construction or handling of proof), which suggest a certain method or approach. Examples are using an $\varepsilon/2$-approach when dealing with ε-δ-continuity or knowing that a task description containing a "Show that X holds for all natural numbers"-formulation, where X stands for an arbitrary mathematical statement that depends on a natural number (e.g., "The expression n^3-n is divisible by 3."), is likely to require a proof by complete induction. Hence, mathematical strategic knowledge refers to domain- and partially content-specific knowledge.

4.1.1.4 Resource 4: Methodological Knowledge

Methodological knowledge (Heinze & Reiss, 2003) refers to knowledge about the functions and the nature of mathematical proof as a concept (see de Villiers, 1990; Hanna & Jahnke, 1996; Reid & Knipping, 2010) and about the acceptance criteria for mathematical proofs within the local mathematical community (see Dawkins & Weber, 2016; Fallis, 2003; Geist et al., 2010; Hanna, 1990; Heinze, 2010; Mariotti, 2006). According to the framework by de Jong and Ferguson-Hessler (1996), methodological knowledge can be interpreted as conceptual knowledge, as it refers to principles that apply within the domain of mathematics.

Although methodological knowledge partly depends on the socio-mathematical norms within the local mathematical community (see section 2.2), three important aspects are often highlighted (Heinze & Reiss, 2003): The *proof scheme*, which relates to the reasoning patterns used for individual inferences within a proof, for example that examples cannot be used in order to prove a statement; the *proof structure*, which refers to the overall structure of a proof, for example leading from the premises to the claim in a linear proof and not allowing circular reasoning; the *chain of conclusions* [12], which relates to the warrants and backings allowed in a proof, that is only prior accepted knowledge and already proven statements are allowed.

So far, several studies underline the importance of methodological knowledge for mathematical argumentation and proof skills (e.g., Healy & Hoyles, 2000; Heinze & Reiss, 2003; Ufer, Heinze, Kuntze, et al., 2009), each time using participants' performance on proof validation tasks as a measure of their methodological knowledge.

[12] Both, „chain of conclusions" (e.g., Heinze & Reiss, 2003) and "logical chain" (e.g., Kuntze, 2008) have been used in prior research to denote this aspect.

4.1.1.5 Resource 5: Problem-solving Skills & Heuristics

A problem is usually defined as a task, in which the problem solver is "not aware of any obvious solution method" (Mayer & Wittrock, 2006, p. 287) and which therefore represents a barrier to its solution[13] (Dörner, 1979, p. 10). Based on this definition, problem solving is understood as an act of cognitive processing, directed at transforming a given situation into a specified goal situation (Mayer & Wittrock, 2006, p. 287). *Problem-solving skills*, that is the disposition to solve problems successfully over a range of domains, are often conceived as a complex cognitive skill (e.g., Schoenfeld, 1985), having various underlying resources on its own. In comparison to mathematical argumentation and proof skills, the resources, as well as the overall skill, are seen as domain-general. In particular, problem-solving skills in general do not require any of the domain-specific resources mentioned before. As problem solving is usually a heuristic process, based on preliminary decisions for strategies that do not necessarily lead to a solution, problem-solving heuristics are often mentioned as a major resource. These comprise strategies, methods, and rules, which can guide problem-solving processes, yet do not guarantee success (e.g., Abel, 2003; Polya, 1945). Examples are means-end analysis, drawing a sketch, or working backwards (see Polya, 1945, p. 69 f.; Schoenfeld, 1985, p. 44 f.). Heuristics are mostly conceived as domain-general, although some conceptions (e.g., Chinnappan & Lawson, 1996; Koichu, Berman, & Moore, 2007) also include domain-specific strategies. To clearly distinguish heuristics and mathematical strategic knowledge, we only regard domain-general heuristics.
A classic example of a domain-general problem is the *three jug problem*. Here, three jugs of a specified size are given, each holding a specified amount of liquid, and the problem solver is asked to obtain a desired amount of liquid in one of the jugs by filling them into each other.

4.1.1.6 Resource 6: Metacognitive Awareness

Based on the work of Flavell (1979, p. 906) metacognition is defined as "knowledge and cognition about [one's own] cognitive phenomena". As cognition includes the encoding, memorizing, and recalling of information, metacognition can be understood as the skill to monitor and regulate these cognitive processes (Schraw et al., 2006). Metacognition is supposed to be important during problem solving (e.g., Schoenfeld, 1987) to monitor one's progress and direct oneself in a (hopefully) successful direction towards the solution of the problem (see also section 2.4.1). Therefore, it is also relevant for handling mathematical argumentation and proof and is suggested as another domain-general resource, also underlying problem-solving skills (e.g., Schoenfeld, 1985). Since metacognition itself is not directly accessible for researchers and usually measured indirectly using students' self-reports on metacognitive knowledge and behavior (Schraw, 1998; Schraw & Dennison, 1994), it is included as "metacognitive awareness", conceptualized as the degree one is aware of one's own metacognition, in this project.

4.1.1.7 Resource 7: Conditional Reasoning Skills

Conditional statements are if-then rules connecting a premises to a conclusion, for example, "If I train hard enough, I can become a Jedi". Conditional reasoning skills thus refer to handling such conditional statements, that is to draw correct inferences and to judge the truth of derived statements, using for example *modus ponens* or *modus tollens* (e.g., Evans, Clibbens, & Rood, 1995; Inglis & Simpson, 2008; Johnson-Laird, 2000; Johnson-Laird & Byrne, 2002). Given the

[13] According to this definition, the status of a task as a problem is not intrinsic to the task, but depends on the person (problem solver) and therefore is interindividually different. For instance, the example given below will be a problem for almost all persons who did not encounter the task before, but after solving a similar problem, the task may not represent a problem anymore.

conditional statement above, an according task is: “Given Yoda is a Jedi, is it true that he trained hard?” Here, the statement “Yoda trained hard” does not follow (necessarily) from the rule above, given that “Yoda is a Jedi”, and therefore the conclusion is wrong (affirmation of the consequent).
Conditional reasoning skills are often connected to deductive reasoning skills in general (e.g., Attridge, 2013; Attridge & Inglis, 2013; Chinnappan et al., 2012; Evans et al., 1995; Evans, Newstead, & Byrne, 1993) and are therefore supposed to be underlying students’ argumentation and proof skills (e.g., Epp, 2003). Conditional reasoning skills are assumed to be domain-general.

4.1.1.8 Resource 8: Beliefs

Mathematics-related beliefs are mostly defined as “implicitly and explicitly held subjective conceptions about mathematics education, the self as a mathematician, and the social context” (De Corte et al., 2000, pp. 689-690; Op ’t Eynde et al., 2002, p. 14). Although there are some ambiguities in the definition and measurement of beliefs (e.g., Törner, 2002), they are awarded a key role in many mathematics related processes such as learning, handling mathematical problems, or handling mathematical argumentation and proof (e.g., Roesken, Pepin, & Toerner, 2011; Schoenfeld, 1985) and are often separated from affective characteristics (e.g., Hannula et al., 2016) as some research traditions emphasize their cognitive rather than affective aspects (see Furinghetti & Pehkonen, 2002, pp. 40-41). Their influence is supposed to be mediated by cognitive as well as conative and affective processes (Op ’t Eynde et al., 2002; Schoenfeld, 2010).

4.1.1.9 Resource 9: Affective Characteristics

Affect regarding mathematics is conceived as influencing success and failure in mathematics in general (e.g., Di Martino & Zan, 2011; Hannula, 2006; Pekrun, 1992) as well as in handling mathematical argumentation and proof in specific. A common example is mathematics anxiety (e.g., Hembree, 1990). Besides beliefs, which are at conceived as at least partially affective[14], Hannula et al. (2016) name motivation and emotions as two broad categories of affect, which are often divided into rapidly changing and continuously fluctuating *state* variables, and relatively stable *trait* variables. Among state variables, academic achievement emotions of students (Pekrun, Goetz, Titz, & Perry, 2002) such as joy, hope, pride, or anxiety have been highlighted as impacting students’ performance, based on the control-value theory (Pekrun, 2006).

4.2 Guiding Research Questions

The overall aim of our MIMAPS project is to find effective means to adequately assess (university students’) mathematical argumentation and proof skills and subsequently to support the acquisition and learning of mathematical argumentation and proof skills. In particular, we are interested in the individual cognitive resources underlying students’ mathematical argumentation and proof skills and their use for research and education.
As the initial step for our project, we created the research framework just outlined, which combines three main perspectives on mathematical argumentation and proof skills and therefore allows the integration of prior research findings from different perspectives into one single framework. It allows to explore the connections between different aspects of mathematical argumentation and proof skills and acquire a comprehensive view of mathematical argumentation and proof skills.

[14] As beliefs are conceived as in parts cognitive and affective, depending on the research tradition (see Furinghetti & Pehkonen, 2002, pp. 40-41), we chose to include beliefs and affective characteristics as two separate resources for this project.

4.2.1 Reviewing Current Research

Based on the newly-created research framework, the first study of the project is a descriptive literature review (see section 5.1) that contributes to a comprehensive view of mathematical argumentation and proof skills by reviewing prior research on mathematical argumentation and proof skills and by examining closer the different aspects focused within current research, as well as the combinations of aspects.

RQ1 Which resources, processes, and situations in the context of mathematical argumentation and proof skills are currently addressed by mathematics education research? Is there research focusing on the individual aspects of mathematical argumentation and proof skills in a comprehensive way? Which combinations of aspects have been examined?

Here, it was of special interest to classify research regarding the different resources, processes, and situations they focused on. It was desired to determine whether research rather addressed single sub-aspects within one of the three aspects, that is for example solely examining beliefs as a single resource, or if research also addressed multiple resources, processes, or situations. Finally, research on relations between the different aspects, that is, for example, studies examining the influence of problem-solving skills on proof validation, were of major interest as they allow to shed a comprehensive light on mathematical argumentation and proof skills, combining the aspects of the framework and highlighting the relations between the aspects of the framework currently examined in mathematics education research.

4.2.2 The Impact of Students' Individual Resources in Different Situations

To obtain a deeper understanding of mathematical argumentation and proof skills in terms of the underlying cognitive resources, our second study (see section 5.2) assessed the influence of several potential resources included in the framework on first-year university students' performance in handling mathematical argumentation and proof skills in different situations. Research was guided by the question:

RQ2 What is the relative influence of the potential individual resources underlying students' mathematical argumentation and proof skills on their performance in proof construction and validation? Can differences regarding the influence of domain-specific and domain-general resources be observed?

To analyze the potential different influence of the underlying individual resources in different situations, proof construction and proof validation were chosen, as they are predominant both within university students' academic life and mathematics education research (Mejía-Ramos & Inglis, 2009a, 2009b). Furthermore, we only included *cognitive* resources outlined in our research framework. This was done for methodological reasons because assessing all nine resources of the framework would have resulted in a high testing load for participants. Also, the approach to first focus on cognitive resources is a common approach used to obtain an initial picture of their influence (e.g., Chinnappan et al., 2012; Herppich et al., 2017; Klieme & Leutner, 2006; Ufer et al., 2008). The approach is further underpinned by the fact that non-cognitive resources are often conceived to modulate respectively moderate the cognitive resources (e.g., Herppich et al., 2017; Op 't Eynde et al., 2002; Schoenfeld, 2010), thus having no or only minor direct effects and potentially to vary heavily across students (see Furinghetti & Morselli, 2009), therefore requiring a very nuanced analysis. Furthermore, also methodological knowledge was excluded because it has been so far mostly assessed indirectly, using the performance in proof validation as a measure (e.g., Heinze & Reiss, 2003; Ufer, Heinze, Kuntze, et al., 2009). Since proof validation was studied as a part of mathematical argumentation and proof skills, we did not include methodological knowledge as a separate resource.

The resources included in the study are shown in Figure 28. Based on prior research findings, we assumed that all these resources have a non-negative relation to students' performance in proof construction and proof validation. For students' performance on proof validation, we expected conceptual mathematical knowledge as well as conditional reasoning skills as two important resources, both needed for checking the individual facts and inferences within the given proofs. For proof construction, we also assumed an influence of conditional reasoning skills, as constructing a mathematical proof entails the construction of a deductive chain of arguments. More importantly we expected an influence of conceptual and procedural mathematical knowledge as well as problem-solving skills as suggested by prior research (Chinnappan et al., 2012; Ufer et al., 2008).

Besides analyzing the influence of the underlying resources in the situations of proof construction and proof validation, the study also aimed to advance the systematization of the relationship between proof validation and proof construction, as their conceptual status as a) two independent skills, b) two subskills of overall mathematical argumentation and proof skills, or c) two situations requiring the same skill, is still unclear (e.g., A. Selden & Selden, 2015a)[15]. Based on the prior research findings discussed in section 3.4.4 that highlight the relation between both and conceptualize proof validation an important process during proof construction, we also addressed the question:

RQ3 How do students' proof validation skills relate to their proof construction skills? Can their proof validation skills add to the explanation of their proof construction skills beyond the included individual resources?

To address this question, we first examined the correlation between proof construction and proof validation skills. Here, we expected a positive, weak correlation based on prior findings (Ufer, Heinze, Kuntze, et al., 2009). Further, we examined the impact of students' proof validation skills when included as another predictor of students' proof construction skills in the statistical models employed. In case it adds to the explanation of students' proof construction skills beyond the included individual resources, this can be interpreted as a sign that proof validation skills themselves adds to proof construction skills or that yet another resource, which was not included in this analysis (e.g., methodological knowledge), is jointly underlying both skills. If alternatively, students' proof validation skills correlate with students' proof construction skills but do not add to its explanation beyond the included individual resources, the correlation between both skills has to be interpreted as an artefact of the included individual resources and that no signs for other shared resources can be found.

[15] Within this project and the underlying research framework, we have chosen perspective c). Still, for reasons of simplicity we will also use the terms *proof validation skills* and *proof construction skills* to denote students' skills in handling argumentation and proof tasks in the according situations.

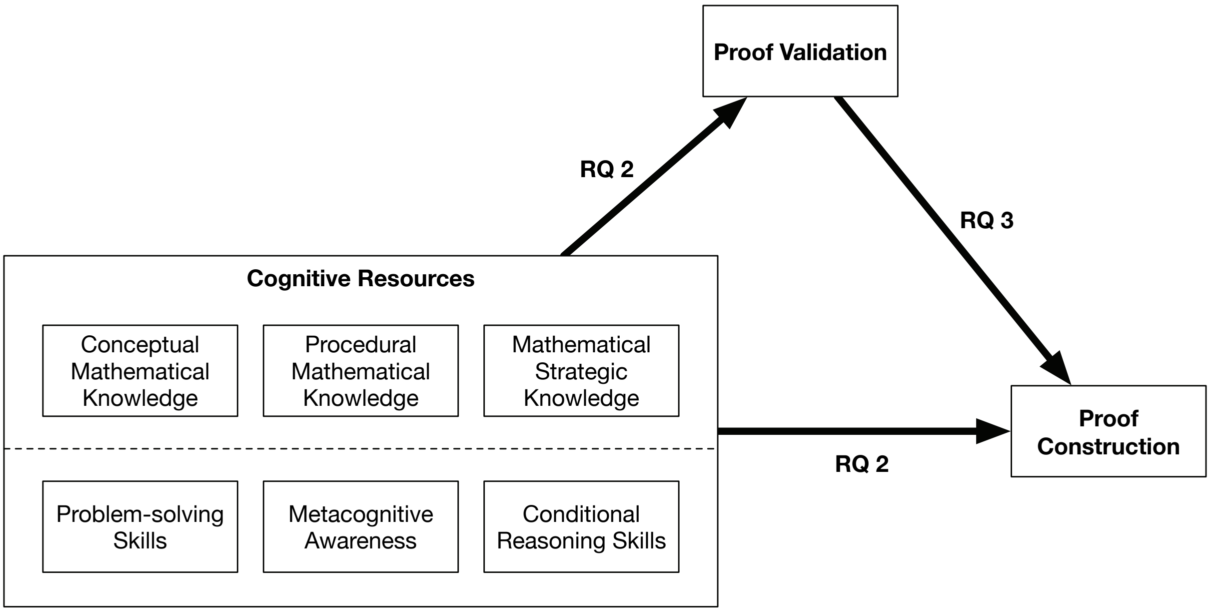


Figure 28. Relation between cognitive resources, proof validation, and proof construction skills examined in the correlational study.

4.2.3 Supporting Mathematical Argumentation and Proof Skills Using Underlying Resources

The third study (see section 5.3) of our project examined the educational implications of conceptualizing mathematical argumentation skills as a complex cognitive skill with several underlying resources.

Based on research from instructional design, especially the part-task / whole-task debate (see section 3.5), two approaches for supporting the resources underlying students' mathematical argumentation and proof skills arise: A *one-by-one* approach, which focuses on and tries to support each resource at a time, and a *simultaneous* approach, which focuses on and tries to support the resources at the same time. Thus, the third study aimed to answer the following question:

RQ4 What are the differences between two instructional approaches that aim at supporting the resources individually one-by-one or simultaneously in terms of students' acquisition of individual resources as well as overall mathematical argumentation and proof skills?

To address this question, four cognitive resources of mathematical argumentation and proof skills were included in a training concept (mathematical knowledge[16], mathematical strategic knowledge, problem-solving skills, and methodological knowledge). Affective characteristics and beliefs were again excluded, this time with the additional reason that several studies had underlined that affective (trait) aspects as well as beliefs regarding mathematics are relatively stable and difficult to change (see Kloosterman, Raymond, & Emenaker, 1996). Further, the number of included, potential resources had to be limited to four in order to decrease the complexity and length of the intervention (both for students and the design of the study in general).

[16] Here, mathematical knowledge refers to conceptual and procedural knowledge, which were not distinguished within the intervention for educational purposes as students might have conceived the difference as problematic.

We hypothesized that the one-by-one approach would be superior regarding learning gains for students' *resources* underlying mathematical argumentation and proof skills. We assumed that students profit from covering the resources individually throughout the intervention, giving them enough room to focus on each one, and having a superior instructional clarity. However, we hypothesized the simultaneous approach would yield better learning gains for students' *overall mathematical argumentation and proof skills*. We expected that students are better able to integrate the resources into overarching strategies to approach proof problems by that approach, as the resources are used concurrently throughout the intervention and students therefore are better able to purposefully apply them when handling mathematical proofs. Further, this hypothesis is underlined by the findings of the part-task / whole-task debate that regard whole-task approaches as superior for the learning of a complex cognitive skill.

5 Studies

5.1 Research on Mathematical Argumentation and Proof Within PME - Results From a Descriptive Review

5.1.1 Introduction

Mathematics[17] is a proving science (Heintz, 2000). As such handling mathematical argumentation and proof represent central mathematical activities (Hanna & Jahnke, 1996; Jahnke & Ufer, 2015; Ubuz, Dincer, & Bulbul, 2012). Many standard documents for secondary schools worldwide put forward mathematical argumentation and proof skills as a central goal of mathematics learning (e.g., Common Core State Standards Initiative, 2010; KMK, 2012; National Council of Teachers of Mathematics, 2000), which becomes even more important at the university level, where mathematical proof is introduced as *the central* method of mathematics as a scientific discipline. Accordingly, argumentation and proof also represent a major line of research in mathematics education and have been approached from a variety of perspectives within the last decades (e.g., Hanna, 1990; Mariotti, 2006).

Today, it is widely agreed that mathematical argumentation and proof skills comprise a complex cognitive skill (Blömeke et al., 2015; Schoenfeld, 1985; van Merriënboer & Kirschner, 2007) and therefore require several underlying resources, such as knowledge facets, skills, beliefs, or affective characteristics, in order to be successful in handling mathematical argumentation or proof (e.g., Blömeke et al., 2015; Schoenfeld, 1985). At the same time, research increasingly focuses on different situations in the context of argumentation and proof such as reading or presenting arguments and examines specific processes therein more closely.

Although it is obviously desirable to examine mathematical argumentation and proof skills from multiple different perspectives and thereby capture its whole breadth, research on underlying resources, processes, and situations in the context of mathematical argumentation and proof need to be purposefully combined to achieve this. Thus, one aim of research on mathematical argumentation and proof skills must be an increasingly coherent understanding of these facets, ultimately leading to the creation of effective means of support for students' mathematical argumentation and proof skills.

To contribute to this aim, we conducted a descriptive literature review, giving an overview of recent research on mathematical argumentation and proof in secondary and tertiary education. This research method does not create new evidence per se and also does not meta-analytically combine prior evidence, but it allows to structure prior research, to highlight connections, and to point out desiderata, therefore building a basis for future research. To extensively cover the diversity of research perspectives on mathematical argumentation and proof skills in this review, we use a broad conceptualization of mathematical argumentation and proof skills as a foundation, guided by a framework of Blömeke et al. (2015) that is embedded in research on teaching skills and the assessment of complex cognitive skills in higher education. Based on this framework, we structure existing research according to three different perspectives on mathematical argumentation and proof skills: The underlying individual *resources*, the *situations*

[17] A shorter and less detailed report of this study has been presented at the conferences PME 39 and GDM 2015 and published in the respective proceedings volumes (Sommerhoff, Ufer, & Kollar, 2015a, 2015b).

that require the use of mathematical argumentation and proof skills, and the *processes* leading to an observable performance.

5.1.2 Theoretical Framework

5.1.2.1 Argumentation and Proof

To set the stage for the review, we first consider the definition of mathematical argumentation and proof skills used in the sequel. There is a general ambiguity regarding the use of the words *argumentation*, *reasoning*, *proof,* and their verbal forms such as *proving* (e.g., Reid & Knipping, 2010). One major distinction pointed out, for example by Balacheff (1999), is that there are two meanings associated with argumentation within the field of mathematics: On the one hand, mathematical argumentation can be characterized as a social-discursive activity aimed at convincing a listener or group of listeners. This view is often highlighted by general educational research (e.g., Nussbaum, 2008; Voss & Van Dyke, 2001) and some works in argumentation theory (e.g., Andriessen, 2009; van Eemeren & Grootendorst, 1999). On the other hand, argumentation can also be seen as an activity that is aimed at the generation, exploration, and validation of (mathematical) conjectures and hypotheses regarding their objective and individual rationality (Kollar et al., 2014; Pedemonte, 2007; Reichersdorfer et al., 2014). This can be done individually or in groups and possibly to persuade others (or oneself), but can also have various other goals (see de Villiers, 1990; Hanna & Jahnke, 1996; Herbst, Miyakawa, & Chazan, 2010).
For this review, the latter view is adopted. Based on this view, mathematical proof is mostly seen as a specific form of mathematical argumentation that is subject to (often implicit, and possibly changing) social norms of the mathematical community (see Dawkins & Weber, 2016; Yackel & Cobb, 1996). Three central, but not exhaustive norms (Heinze & Reiss, 2003) are the sole use of deductive inferences, an appropriate structure of the proof, and the explicit reference to the underlying mathematical theory. This difference between argumentation and proof is characterized by (Pedemonte, 2008, p. 385):

> There is a "structural gap" between argumentation and proof because in argumentation inferences are based on content while in proof they follow a deductive scheme (data, claim, and inference rules).

Accordingly, *mathematical argumentation and proof skills* refer to students' skills in handling mathematical argumentations and proofs, as members of a mathematical community.
Based on this conceptualization of proof, we limit the scope of our review on research on mathematical argumentation and proof in secondary and tertiary education, as *proof* is usually introduced in secondary school and does not represent a central theme before.

5.1.2.2 Mathematical Argumentation and Proof Skills as a Complex Cognitive Skill

Mathematical argumentation and proof skills are often interpreted as a domain-specific *complex cognitive skill* and, especially in the European context, as a competence (see Klieme & Leutner, 2006; Koeppen et al., 2008; Weinert, 1999). That is, they are conceptualized as a latent cognitive and (partially) affective-motivational disposition underlying a person's performance in certain situations (Koeppen et al., 2008).
In accordance with this view, the framework of Blömeke et al. (2015) highlights three aspects of a general complex cognitive skill: Different *cognitive and affective-motivational dispositions* underlying certain *situation-specific skills* that in turn lead to *observable behavior*.
Here, we adapt this framework by Blömeke et al. (2015) as it combines three primary foci in research on mathematical argumentation and proof skills: The *resources underlying*

mathematical argumentation and proof skills (e.g., Chinnappan et al., 2012; Schoenfeld, 1985; Ufer et al., 2008), the *processes* they are enacted in (e.g., Boero, 1999), and the performance in specific *situations* that may pose different domain-specific demands on students (e.g., Mejía-Ramos & Inglis, 2009a; Mejía-Ramos & Inglis, 2009b).

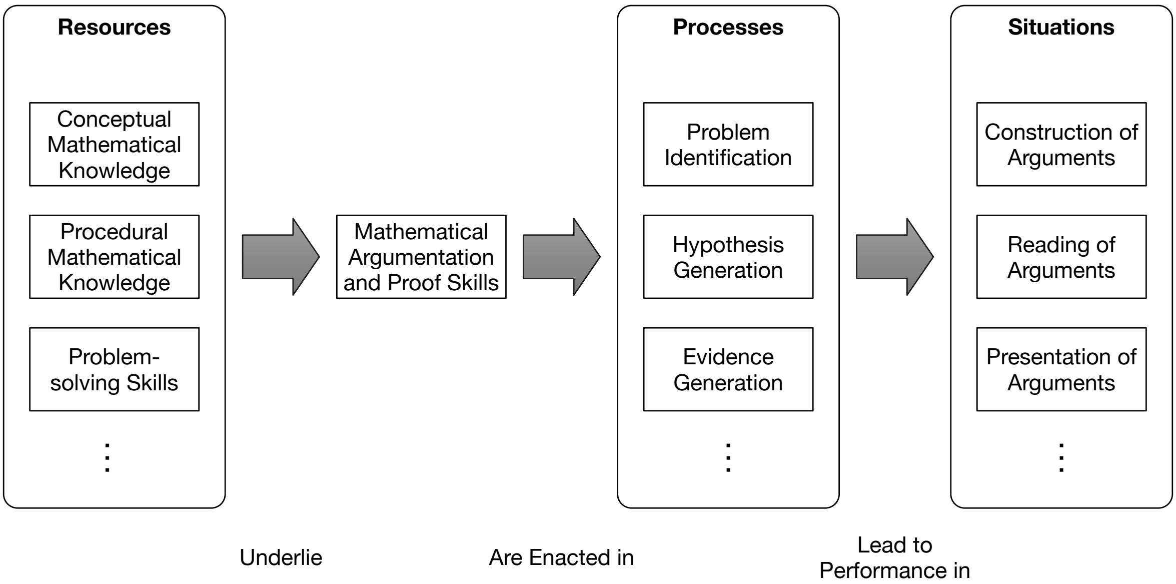


Figure 29. Three aspects of mathematical argumentation and proof skills as a complex cognitive skill; adapted from the framework by Blömeke et al. (2015).

Each of these three aspects of mathematical argumentation and proof skills are elaborated on in the following, using prior frameworks from mathematics education to describe different sub-aspects, resulting in the research framework used in this review (see Figure 29), which is then used to characterize recent research on mathematical argumentation and proof skills with respect to its focus on either of these three aspects, specific sub-aspects, and their combinations.

5.1.2.2.1 Resources Underlying Mathematical Argumentation and Proof Skills

Within the last decades, several resources for mathematical argumentation and proof skills have been proposed, like mathematical content knowledge such as concepts and procedures, methodological knowledge, or problem-solving skills (e.g., Heinze & Reiss, 2003; Schoenfeld, 1985). The proposed resources vary regarding their domain-specificity (see Chinnappan & Lawson, 1996) as well as regarding their nature as knowledge facets, beliefs, and affective characteristics (de Jong & Ferguson-Hessler, 1996; Hannula, 2006).

The resources for mathematical argumentation and proof skills included in this review were partially derived from research on more general activities such as *problem solving* (e.g., Schoenfeld, 1985) or self-*regulated learning* (e.g., De Corte et al., 2000), or were inferred from qualitative studies, as for example mathematical strategic knowledge (e.g., Weber, 2001). Up to now, no exhaustive list of resources of mathematical argumentation and proof skills is available and there is only little empirical evidence available regarding the predicitivity of each proposed resource for students' performance in situations requiring mathematical argumentation and proof skills. To date, mostly either qualitative studies have been conducted, or the quantitative studies have focused only on single resources, therefore neither accounting for the possibly of confounded variables nor for interactions between the variables, both of which may alter the results of the hitherto existing studies. Here, only first studies (e.g., Chinnappan et al., 2012; Ufer et al., 2008), have been conducted, examining multiple resources at once and thereby contributing to more reliable estimates for the individual predictivity of the resources.

Combining various central frameworks and studies (Chinnappan et al., 2012; De Corte et al., 2000; Leder, Pehkonen, & Törner, 2002; Schoenfeld, 1985; Ufer et al., 2008; Weber, 2001), we identified six resources that were deemed the most relevant resources for this review:
Mathematical content knowledge (MCK) consists of knowledge of those mathematical concepts that form the context or basis for the argumentation processes. This comprises conceptual knowledge of definitions, facts, and properties, as well as procedural knowledge, such as technical, routine aspects and procedures (Hiebert, 1986; Rittle-Johnson et al., 2015).
Mathematical strategic knowledge (MSK) refers to knowledge about cues and hints within mathematical tasks and problems that indicate concepts, methods, or representation systems that can be used productively for the given task (Weber, 2001).
Methodological knowledge (MK) is knowledge of the nature and the functions of the concept mathematical proof. It furthermore contains knowledge about the local and global acceptance criteria for a valid proof (Healy & Hoyles, 2000; Heinze & Reiss, 2003; A. J. Stylianides, 2007).
Problem-solving skills (PSS) consist of general problem-solving skills and knowledge about general heuristics, such as means-end-analysis (e.g., Chinnappan et al., 2012; Polya, 1945; Schoenfeld, 1985), which are not limited to the domain of mathematics.
Finally, there are *beliefs (BE)* about mathematics education, the self as a mathematician, and the social context (e.g., Leder et al., 2002; Op 't Eynde et al., 2002, p. 14; Schoenfeld, 2010) as well as *affective characteristics (AC)* such as emotions and motivation towards mathematics (e.g., Hannula, 2006).

5.1.2.2.2 Processes of Mathematical Argumentation – The Epistemic Activities Framework

Besides the resources underlying mathematical argumentation and proof skills, special emphasis has to be put on the specific processes that individuals apply to master situations that pose argumentation and proof demands. These processes have been described from various perspectives, for example as processes during problem solving (Polya, 1945), the processes of (expert) mathematicians when constructing proofs (Boero, 1999; Kirsten, 2017), or as the activities involved in general scientific reasoning practices (Fischer, Kollar, et al., 2014).
The latter framework, proposed by Fischer, Kollar, et al. (2014), focuses on *scientific reasoning and argumentation* from an interdisciplinary point of view. Research on these processes is often seen as initiated by findings by Inhelder et al. (1958), has advanced throughout the last decades (see Kuhn, 2002; Sodian & Bullock, 2008; Zimmerman, 2000), and focuses on the activities during processes of scientific discovery and inquiry, such as identifying problems, formulating research questions and hypotheses, generating evidence for these and evaluating it (Fischer, Kollar, et al., 2014; Kuhn, 2002). Accordingly, mathematical argumentation and proof skills can be seen as a domain-specific instantiation of scientific reasoning and argumentation skills and can, therefore, be described using the general process model for scientific reasoning and argumentation by Fischer, Kollar, et al. (2014). The framework is helpful, as it contains more processes and is, therefore, more nuanced than the other models (e.g., Boero, 1999) and allows to describe processes in situations other than proof construction. Finally, it also permits to compare results of the review with other domains, as the same terminology is used.
The framework comprises eight so-called *epistemic activities* (see Table 2 for a description of each and an illustration in the context of students within a geometry classroom) that are sub-processes of scientific reasoning and argumentation processes and can be used to describe the (cognitive[18]) processes during mathematical argumentation. Albeit their linear presentation,

[18] Obviously, the actual cognitive processes cannot be observed. The epistemic activities, however represent observable processes that are suggested to correspond to different cognitive processes and therefore can be used as proxies of these.

they do not need to occur in this specific order, can be iterated, possibly in loops, and not necessarily all of them have to be present in an argumentative process.

Table 2. Overview of epistemic activities by Fischer, Kollar, et al. (2014).

Epistemic Activity	**Description**	**Illustrations**
Problem Identification (PI)	Perceiving a mismatch concerning the explanation of a problem and building a problem representation	Students observe (seemingly) contradicting statements that call for a scientific explanation or clarification.
Questioning (QU)	One or more initial questions are identified	Students ask themselves whether it is possible that both statements are true at the same time.
Hypothesis Generation (HG)	Possible answers to the questions are derived from models, theoretic frameworks, or other sources	Students generate the hypothesis that it can be inferred from statement 1, that statement 2 is false.
Construction and Redesign of Artefacts (CR)	Development of a prototypical object, axiomatic system, or another object used to work on the problem	Students generate a DGS-worksheet[19] to (inductively) check the hypothesis.
Evidence Generation (EG)	Evidence for the hypothesis is generated	Students use deductive inferences to conclude from statement 1 that statement 2 cannot be true.
Evidence Evaluation (EE)	Evaluating evidence according to certain norms	Students check after every step / inference whether these are coherent, support the hypothesis, and lead to the desired goal.
Drawing Conclusions (DC)	Integrating different pieces of evidence, reevaluating the initial claim considering the new evidence	Students reflect upon the generated evidence, connect different parts of the proof, and if necessary reevaluate the initial hypothesis based on the produced evidence.
Communicating and Scrutinizing (CO)	Sharing and discussing argumentations within a community	Students present their arguments / proof to each other and check their validity critically.

[19] DGS-worksheets are digital representations of geometric figures within a dynamic geometry system (e.g., GeoGebra, Cinderella), which allow the dynamic reconfiguration of objects, for example by dragging.

5.1.2.2.3 Situations Involving Mathematical Argumentation and Proof Skills

The third aspect highlighted by Blömeke et al. (2015) is the performance in different *situations* that pose different demands on students. These can be structured using a framework suggested by Mejía-Ramos and Inglis (2009a, 2009b), categorizing different activities in the context of argumentation and proof based on work by Giaquinto (2005) regarding their givens, goals, and products. Mejía-Ramos and Inglis divide argumentative activities associated with mathematical argumentation and proof into the three categories *construction of novel arguments*, *reading arguments* and *presenting arguments*, each with a few sub-categories (see Figure 30). Even though this distinction seems partly similar to the processes described by the epistemic activities, especially since Mejía-Ramos and Inglis talk about "activities", their categorization rather refers to the different overall goals of mathematical argumentation and proof activities and therefore refers to situations, rather than to the (sequence of) sub-processes within these. Out of the three situations not all are (perceived as) equally important in mathematics and mathematics education research: In their review Mejía-Ramos and Inglis (2009a) reveal that a large proportion of research is focused on the construction of novel arguments. Still, Mejía-Ramos and Inglis (2009a) also highlight the importance of the other two situations in educational learning and assessment settings. In recent research, reading arguments is seen as multi-faceted (e.g., A. Selden & Selden, 2015a) and is mostly perceived from a student's perspective (e.g., Alcock, Hodds, Roy, & Inglis, 2015; Hodds et al., 2014; Lin & Yang, 2007; Samkoff & Weber, 2015; A. Selden & Selden, 2003; Yang & Lin, 2008), whereas the scarce research on argument presentation mostly relates to teachers and lecturers (e.g., Fukawa-Connelly, 2014; Lai & Weber, 2014; Roy et al., 2010).

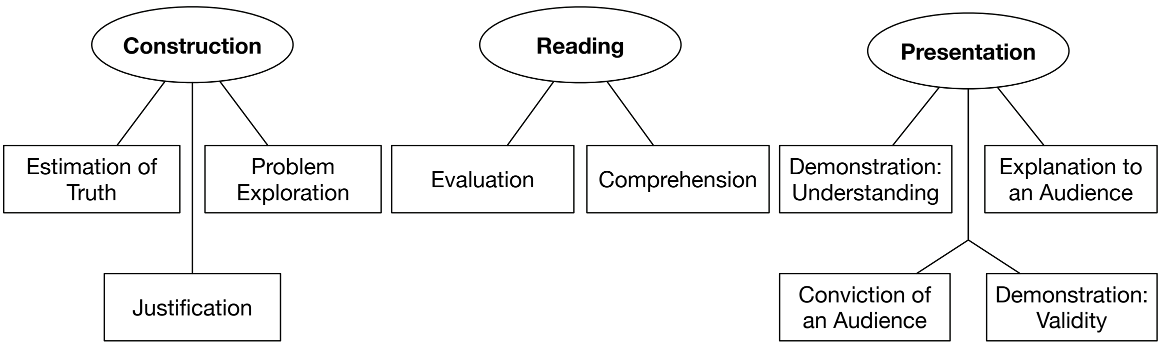


Figure 30. Activities in the context of mathematical argumentation and proof (ovals) and their sub-categories (rectangles) (Mejía-Ramos & Inglis, 2009a).

5.1.3 Aims and Research Questions

Our descriptive literature review sets out to analyze recent research on mathematical argumentation and proof skills in secondary and tertiary education to provide a comprehensive view on this research and thereby support the creation and refinement of a coherent view of mathematical argumentation and proof skills.

To achieve this, we investigated which of the three aspects of mathematical argumentation and proof skills highlighted by our research framework have been examined within mathematics education research between 2010 and 2014. Here, we specifically not only examined the aspects individually but also explored which combinations have been researched and give insights into their connection. By identifying these patterns within research on these aspects, new directions for future research in understanding and supporting mathematical argumentation and proof skills as a complex cognitive skill can be identified.

The review was guided by the following questions:

RQ1 To which extent does research on mathematical argumentation and proof consider the different resources, sub-processes, and situations associated with mathematical argumentation and proof skills?

RQ2 Does research explore connections between the three aspects? Which combinations of resources, processes (epistemic activities), and situations are being considered in research on mathematical argumentation and proof skills? Can research gaps be identified with regard to a comprehensive understanding of mathematical argumentation and proof skills?

As the review is of exploratory nature, no explicit prior hypotheses were made.

5.1.4 Method

5.1.4.1 Literature Selection

As the basis for the review, the proceedings of the International Group for the Psychology of Mathematics Education (IGPME) were chosen, as the IGPME represents one of the largest, international societies for mathematics education research. Within the proceedings, we only considered the research reports (RRs), that is the highest contribution category, which is only accepted after a double-blind peer review by at least two reviewers and has acceptance rates around 50%. Overall, the research reports can be considered as a fair representation of recent, good-quality, international mathematics education research.

Out of the proceedings, the research reports from five years, published between 2010 and 2014, were selected, amounting in a total of 782 research reports.

5.1.4.2 Coding

The coding of the research reports was divided into two rounds: Within the first round, an initial coding regarding the research topic and the focused educational level was performed. The coding was meant to extract the research reports relevant to this review, that is research reports focusing on mathematical argumentation and proof within secondary or tertiary education. The content of the whole research report was used for this coding, not only the abstract.

For the second round of coding, a detailed, theory-based coding scheme was created to categorize the research reports according to the *resources* investigated, the *epistemic activities* studied, and the *situation(s)* examined in the study.

The research foci of each report concerning the three aspects and according sub-aspects were then coded based on the coding scheme, again using the content of the whole research report. That is, those resources, processes, and situations that were related to the research interest of each report were extracted. Any resources, processes, and situations that were not part of the research focus were not included. For example, if participants of a study were talking about or discussing a mathematical problem only for methodological purposes of the study (e.g., as a "thinking aloud" technique) this would neither be coded as the processes of *communication and scrutinizing*, nor as the situation of *presenting arguments*, as none of these were of research interest (see for example Buchbinder & Zaslavsky, 2013).

Resources: For each of the resources, we coded if it was a variable *central* to the research reports (e.g., it was the sole focus of the report), if it was considered as *substantial* (e.g., it was analyzed together with other resources), if it was only *mentioned* (e.g., as a variable to be controlled), or if it did not occur at all. We also included a category for further resources, which had not been included in our list.

Processes: Each of the processes from the epistemic activities framework was coded individually and on different levels. Given the case that a process was part of the research interest of a research report, it was assigned to one of the following four codes, based on the way it was addressed in the research report: It was coded as *normative*, if the process was characterized from a normative point of view, for example when giving a theoretical model for it. It was coded *descriptive*, if the process was characterized in a descriptive way, for example when describing a students' hypothesis generation. Furthermore, it was coded as *problem*, if a specific study addressed the process as a reason for students' problems, which was either revealed as such or for which further evidence was given. Finally, it was coded as *supported*, if the process was explicitly supported or fostered, for example when using scaffolds to support students' hypothesis generation.
Situations: The situations involving mathematical argumentation and proof skills were coded in the categories *argument construction*, *argument reading*, and *argument presentation* where possible. Furthermore, the codes *not explicit* and *multiple* were introduced, when either the situation was not explicated in the report, for example in a theoretical paper, or the research interest concerned multiple situations.

Additionally, we coded information about the research method employed in the research reports and the number of participants in case of an empirical study. We also coded further information, for example relating to the location of the researchers, types of analyses employed, and connections to professional development of teachers, but these will not be part of this article.
As a third step, the results of the descriptive analysis were safeguarded by rereading several examples of research reports from each category to ensure coding quality and to gain a further qualitative insight into the research reports.

5.1.4.3 Coding Quality

The coding scheme described above was refined in several rounds of coding training by two coders. After these rounds, the inter-rater reliability for the coding reached a good level with a mean inter-rater reliability of $\kappa_{Mean} = .77$ ($SD = .15$), based on a double coding of 15% of the research reports. Except for the inter-rater reliabilities of the epistemic activities *drawing conclusions* ($\kappa = .56$) and *communicating and scrutinizing* ($\kappa = .46$) all inter-rater reliabilities were acceptable ($\kappa > .64$).

5.1.5 Results

5.1.5.1 Descriptive Results

The first round of coding revealed that out of the total sample of 782 research reports, 532 (68%) articles were situated in secondary (44%) or tertiary (24%) education and 160 (20%) research reports focused on mathematical argumentation and proof (MA&P) (see Table 3).

Table 3. Distribution of research reports regarding the educational level and focus on mathematical argumentation and proof based on the first round of coding.

		Focus: MA&P	Focus: Other Focus	
Educ. Level	Sec./Ter. Level	**129 RRs**	403 RRs	532 RRs
Educ. Level	Other Educ. Level	31 RRs	219 RRs	250 RRs
		160 RRs	622 RRs	782 RRs

Based on these results, 129 (16%) research reports met the inclusion criteria of being situated in secondary and tertiary education *and* focusing on mathematical argumentation and proof. Examining the research methods used within these research reports (see Figure 31; left side), a clear focus on qualitative research (57%) can be seen, followed by quantitative research (26%), mixed methods (11%), and theoretical papers (6%). This is also reflected by the results regarding the number of participants (see Figure 31, right side). Although the number of participants spans the whole range from 0 to 2590, the highest percentage of articles can be found in the range from one to five (23%), respectively in the categories from 1 to 20 participants[20].

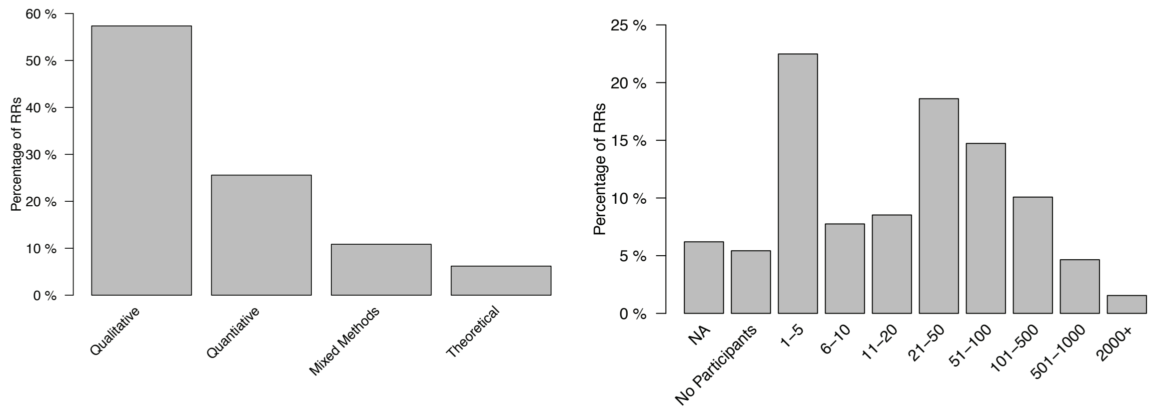


Figure 31. Distribution of research reports regarding research method (left) and number of participants (right).

5.1.5.2 Research on the Three Aspects of Mathematical Argumentation and Proof Skills

Resources

Following our research framework outlined above, we started our analysis with the resources. Here, *mathematical content knowledge* was studied most frequently (47% of the research reports; see Figure 32). Only 24% considered *methodological knowledge* and 18% *problem-solving skills*. *Mathematical strategic knowledge*, *beliefs,* and *affective characteristics* were studied even less frequently (3%, 5%, and 3%, resp.). All in all, only 22% of the research reports considered at least two resources simultaneously (i.e., two or more resources were coded as central or substantial), and over two-thirds of these cases focused on *mathematical content knowledge* in combination with any one other resource.

[20] Please notice that the categories used to analyze the number of participants are not of equal size in order to allow a finer analysis and capture different kinds of studies. However, the differing sizes also lead to the low number of studies with *six to ten* and *eleven to 20* students as compared to the higher categories.

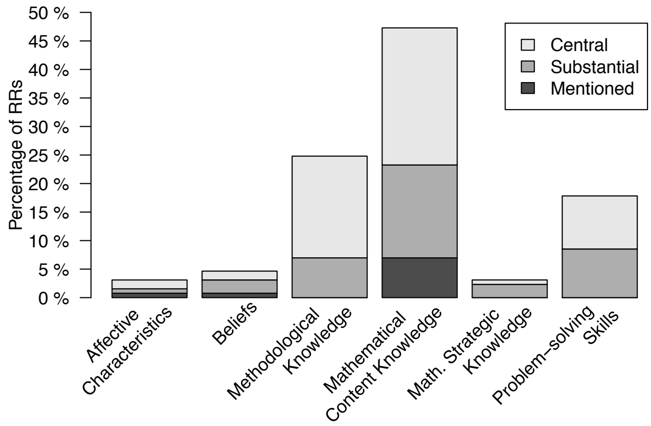


Figure 32. Resources focused on within the research reports.

Situations

Regarding the situations involving mathematical argumentation and proof skills (see Figure 33, left side), almost 60% of the research reports focused on *argument construction*, whereas only 7% examined *argument reading* and 1% *argument presentation*. The number of reports including two or more situations is also low with 7%. Almost a third of the research reports (29%) could not be associated with one of these three activities, for example, because the report was theoretical or did not specify any activity.

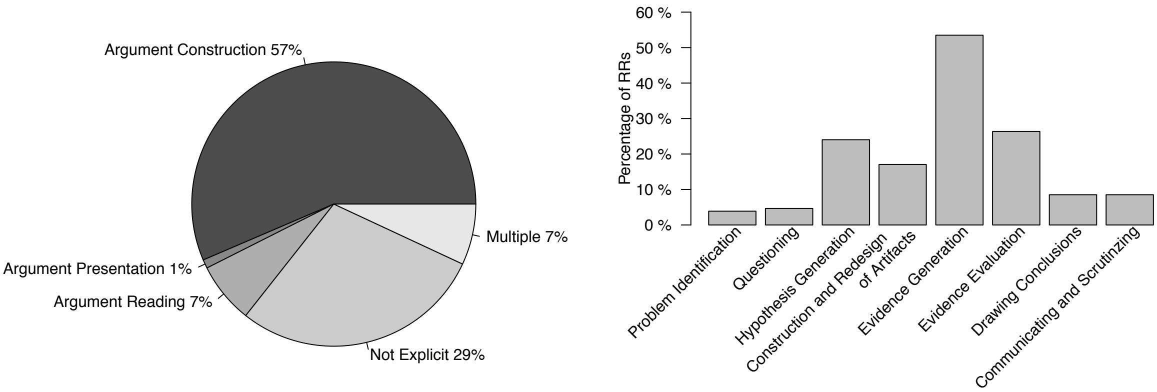


Figure 33. Situations (left) and epistemic activities (right) focused in the research reports.

Processes

In line with this focus on argument construction as the most important situation focused on in the research reports, *evidence generation* was the most frequently studied epistemic activity (53%; see Figure 33, right side), followed by *evidence evaluation* (26%) and *hypothesis generation* (24%). Nevertheless, all epistemic activities were studied at least in some form in some research report.

Examining the different perspectives on epistemic activities, data reveal that half of them (*problem identification*, *questioning*, *construction and redesign of artefacts,* and *drawing conclusions*) were only characterized descriptively within the research reports. *Hypothesis generation* and *generating evidence* were described normatively once, whereas communicating and scrutinizing were described normatively 10 times. Solely *evidence generation* and *evidence evaluation* were characterized as problems (3 respectively 2 times) and were object of explicit support (3 times each) within the research reports.

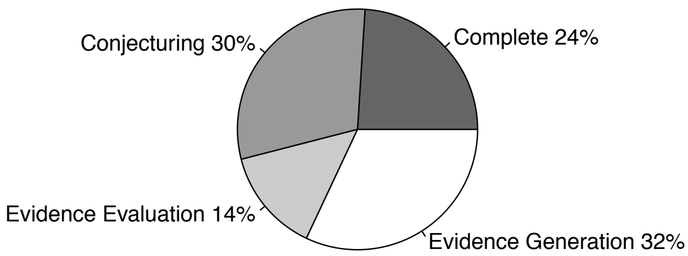


Figure 34. Percentage of research reports belonging to the four clusters of processes.

A qualitative analysis of those research reports that are focusing on at least one epistemic activity (96 of 129 research reports), revealed four clusters of research reports (see Figure 34). This finding could also be supported by a subsequent TwoStep cluster analysis using Schwarz's Bayesian Criterion and a Log-Likelihood distance measure based on the occurrence of the epistemic activities as criteria for similarity.

From the resulting clusters, two focus on one epistemic activity only, whereas the other two focus on multiple epistemic activities concurrently. Out of the four clusters the *evidence generation cluster* that solely focuses on evidence generation, constitutes the largest cluster with 32% of the 96 research reports included in this particular analysis. A representative of this cluster is a research report on unjustified assumptions in geometry proofs, which analyzed students' written geometry proofs regarding these assumptions (Dreyfus & Dvora, 2011). The second largest cluster with 30% of the research reports is the *conjecturing cluster* that focuses on the activities of hypothesis generation, construction and redesign of artefacts, and evidence generation (see Figure 35, upper part). A representative of this cluster is a qualitative study examining how students use dynamic geometry software (i.e., GeoGebra) to obtain solutions for a geometry task for example by systematically dragging points (Jacinto & Carreira, 2013). The third biggest cluster with 24% of the research reports is the *complete process cluster* that incorporates research reports focusing on a broad range of epistemic activities at once (see Figure 35, lower part). A representative of this cluster is a research report on the role of dynamic geometry on the process of exploration, conjecturing, and proving geometrical problems (Samper, Camargo, Perry, & Molina, 2012). Finally, the smallest cluster with 14% of the research reports is the *evidence evaluation* cluster that focuses solely on the epistemic activity of evidence evaluation. A representative of this cluster is an eye-tracking study examining the role of pictures while reading proofs (Beitlich et al., 2014).

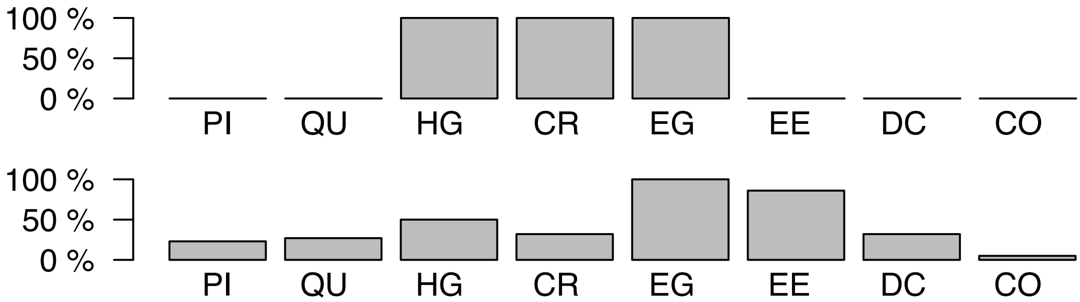


Figure 35. Percentage of research reports within the conjecturing cluster (upper part) and complete cluster (lower part) focusing on each epistemic activity.

Besides the differences regarding their focus within the argumentation process, all four clusters also differ regarding their mean sample sizes and the employed research methods. The mean sample size in the *evidence generation cluster* is 85, whereas the other clusters have mean sample sizes of 50 and below, with the *evidence evaluation cluster* having the smallest mean sample size of 32. All four clusters mostly contain qualitative research reports. Here, both clusters examining multiple epistemic activities (i.e., *complete* and *conjecturing* clusters) comprise an especially high percentage of qualitative research with 77% respectively 79%, whereas the percentage of quantitative studies is higher in those clusters limited to one epistemic activity (38% *evidence evaluation*, 29% *evidence generation*).

5.1.5.3 Connections Between the Three Aspects

To identify patterns within recent research relating to the three aspects of mathematical argumentation and proof skills, we extracted those research reports that simultaneously focused on two of these aspects, respectively certain sub-aspects, and used bubble charts to visualize the according combinations (see Figures 36, 37, and 38). Within the bubble charts, each combination of sub-aspects from both investigated aspects is represented by a circle whose area is proportional to the number of research reports that fall within this combination of categories.

Processes and Situations

Data revealed a close connection between the situation examined and the processes investigated in the context of mathematical argumentation and proof skills in the reviewed research reports (see Figure 36). That is, research reports with a focus on *argument construction* predominantly studied the activities of hypothesis and evidence generation (see Figure 36, lowest line), but also occurrences of all other epistemic activities can be found in this situation. In contrast, the research reports on *argument reading* focused exclusively on evidence evaluation (see Figure 36, second lowest line), and the few research reports on *argument presentation* focused on evidence evaluation and communicating and scrutinizing. Finally, those studies considering multiple situations focused relatively evenly on most processes except for drawing conclusions.

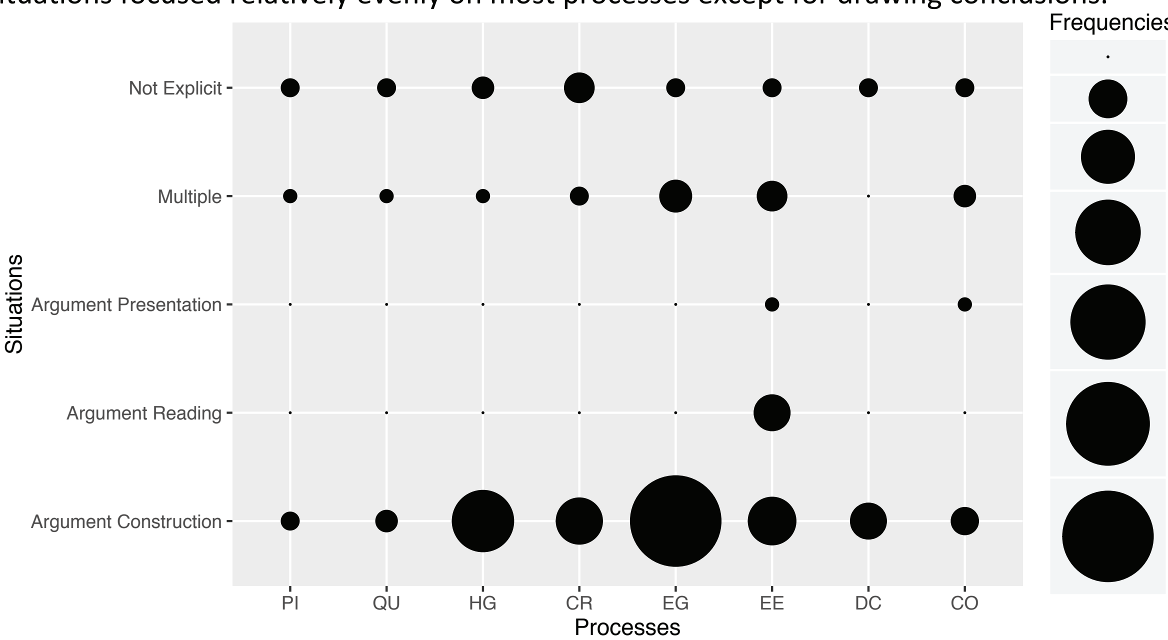


Figure 36. Bubble chart of connections between processes and situations.

Processes and Resources

Examining Figure 37, it can be observed that mathematical content knowledge as well as methodological knowledge are examined in connection with all sub-processes of mathematical argumentation and proof skills within the examined research reports, whereas most other resources are limited to few processes. Here, especially the omission of the two first processes (problem identification and questioning) in connection with the resource problem-solving skills is remarkable.

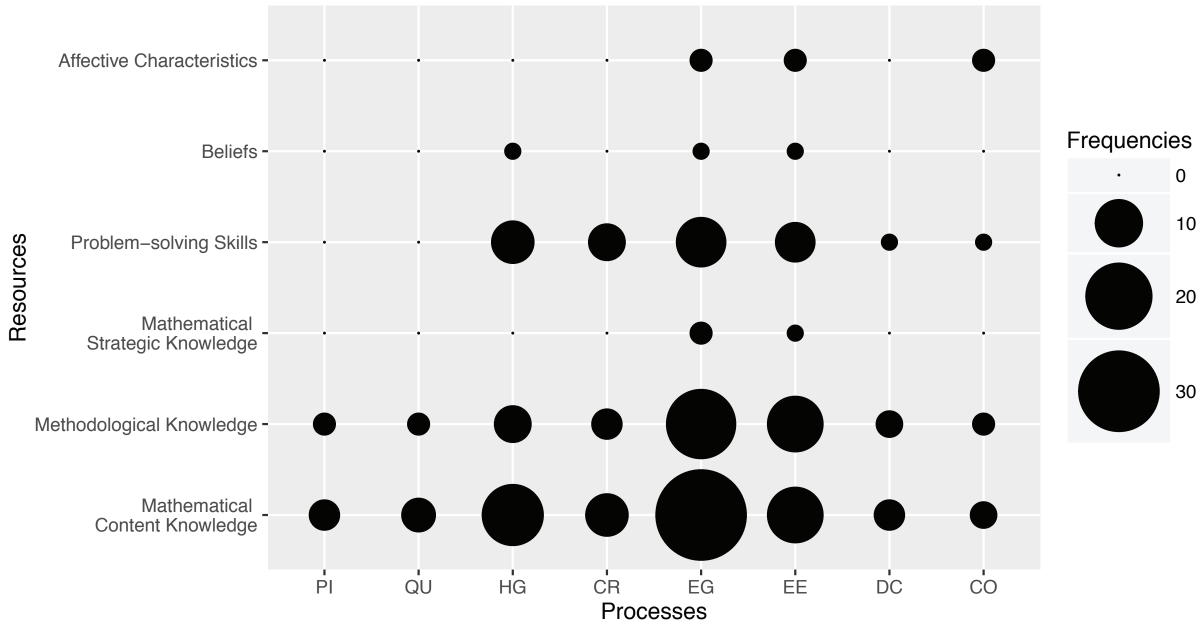


Figure 37. Bubble chart of connections between processes and resources.

Resources and Situations

Data on the connections between resources and situations reveals that *argument construction* is perceived as closely related to mathematical content knowledge, methodological knowledge, and problem-solving skills, yet mathematical strategic knowledge, beliefs, and affective characteristics are rarely among the focused resources (see Figure 38). As there are only few studies examining argument reading and argument presentation, little can be said about the resources included within that research. Still, for *argument reading*, a picture similar to argument construction emerges, however the focus is more on methodological knowledge than on content knowledge and problem-solving skills. A*rgument presentation* is only linked to methodological knowledge and affective characteristics, and none of the other resources is being regarded in this situation.

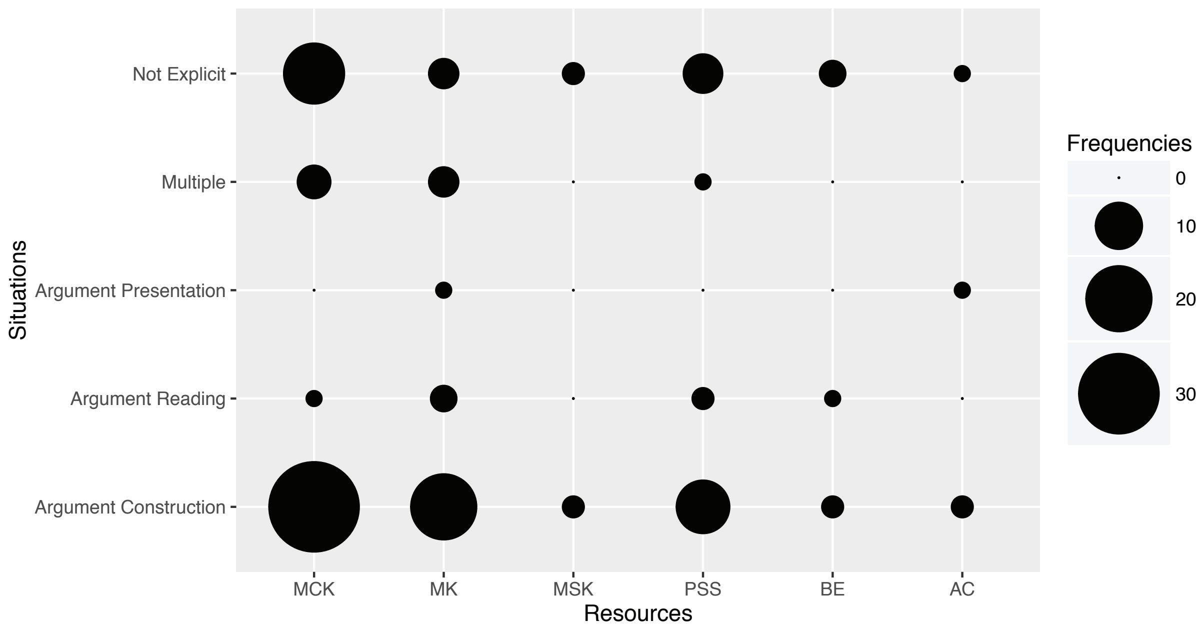


Figure 38. Bubble chart of connections between resources and situations.

5.1.6 Discussion

The aim of our review was to analyze the inclusion of different *resources*, *processes*, and *situations* as well as their combinations in recent mathematics education research on argumentation and proof skills. The results illuminate to which extent recent research contributes to a comprehensive understanding of mathematical argumentation and proof skills, aiming to uncover gaps in the recent research landscape and to reveal perspectives for further research.

Findings regarding the inclusion of different resources in research show a clear focus on only three: *mathematical content knowledge*, *methodological knowledge*, and *problem-solving skills*. As mathematical content knowledge and problem-solving skills are widely discussed in the relevant literature and their influence has also been empirically underlined (e.g., Chinnappan et al., 2012; Ufer et al., 2008), this comes as no surprise. Still, affective characteristics and beliefs have been highlighted as important resources for a long time (e.g., Furinghetti & Morselli, 2009; Leder et al., 2002; Morris, 2007; Schoenfeld, 1985; Stylianou et al., 2015; Weber & Mejia-ramos, 2014), thus the low number of research reports focusing on these is somewhat puzzling. Reasons may be related to issues regarding their conceptualization and assessment (e.g., Törner, 2002) or the fact that affective characteristics (especially trait aspects) and beliefs have partially proven to be hard to change and therefore may be regarded less (Kloosterman et al., 1996). Furthermore, the low number of published research reports is likely to be an artefact of publication bias (see Kühberger, Fritz, & Scherndl, 2014) as beliefs and affective characteristics in general are often perceived as rather modulating the effects of cognitive resources and not having larger direct effects (e.g., Herppich et al., 2017; Op 't Eynde et al., 2002; Schoenfeld, 2010). Thus, such research showing low or no effects may not have been accepted for publication.

The findings further emphasize that most studies consider only single resources of mathematical argumentation and proof skills without accounting for other resources at the same time and therefore not controlling for possible confounded variables or interactions between different resources. Such focused analyses on individual resources are necessary as a first approach to understand complex cognitive skills and their resources. Nevertheless, they bear the potential of overlooking important effects arising from the interplay of multiple resources such as confounding variables or interaction effects, as overall mathematical argumentation and proof skills require the coordination of multiple resources within different processes and different situations. Here, further studies including multiple resources and comparing the relative influence of these resources on mathematical argumentation and proof skills would be valuable. Following Koedinger (1998), such knowledge is particularly important to develop a model of mathematical argumentation and proof skills as well as its resources, which would be a key step towards creating effective learning environments.

The results regarding the processes in focus of recent research show a strong concentration on four epistemic activities, especially *evidence generation* and *evidence evaluation*. Based on standard tasks in the context of proof at university, usually starting with "Prove, that" (Iannone & Inglis, 2010; A. J. Stylianides & Stylianides, 2006, p. 205), this is not surprising. In this context, students neither need to identify a problem, nor develop a question or a hypothesis, as the statement to prove is already known to be true (Heinze, Cheng, Ufer, Lin, & Reiss, 2008, p. 451). Yet, especially instruction in secondary school also offers different situations where students engage in mathematical inquiry and conjecturing (e.g., Koedinger, 1998; Lin, Yang, Lee, Tabach, & Stylianides, 2012), work with (mathematically) authentic problem situations (e.g., Kramarski, Mevarech, & Arami, 2002; Weiss, Herbst, & Chen, 2009), and work collaboratively, discussing and arguing about the topics (e.g., Chen & Chiu, 2015; Chinn & Clark, 2013; Reichersdorfer et al., 2012). At least in these situations, the other processes should be of high importance, which so far is only partially reflected in mathematics education theory (Boero, 1999) and, according to our results as well as those by Mejía-Ramos and Inglis (2009a), also only scarcely in research. As these activities gain importance in curricula worldwide (e.g., KMK, 2012; National Council of Teachers of Mathematics, 2000), more emphasis should be put here in future research.

The analysis of the research reports according to the processes they focus on revealed that only about 50% of the research reports consider more than one of the epistemic activities, reflecting that research on the whole argumentative process is scarce. Here, a cluster analysis revealed more detailed findings, highlighting the existence of four clusters regarding the researched processes: The *complete cluster*, whose reports mainly examine the whole argumentative process (as identified by the epistemic activities), made up only 24% of the examined reports involving at least one of the processes explicitly. However, with 32% the *conjecturing cluster* represents a substantial amount of research focusing on conjecturing and hypotheses generation, a focus that is important for example for inquiry learning, mathematical modeling, or argumentative activities beyond proof. Although controlled by a qualitative check of the research reports within each category, we acknowledge, that the validity of a cluster analysis on binary variables can be questioned. Thus, it would be valuable for future research to employ a more detailed coding of the processes, allowing a more meaningful cluster analysis. This was also intended for this review by including different ways of handling the processes in research (normative, descriptive, problem, and supported; see section 5.1.4.2). However, the more elaborate coding scheme resulted in more categories, thereby reducing the number of studies per category. A viable alternative for future research may be the inclusion of further variables, for example different subject groups, distinguishing pupils, students, and professionals.

Examining the results of the literature review regarding the focused situations, the construction of novel arguments appears to be the main situation of contemporary research interest (57% of the included research reports). This matches the results of Mejía-Ramos and Inglis (2009a) perfectly, who found a focus on argument construction in 63% of their sample, but no contribution at all to argument presentation and few regarding the reading of arguments. This strong focus on situations where the construction of novel arguments is the main goal may be one reason why certain epistemic activities, respectively their combinations, are studied in more detail than others. However, already Mejía-Ramos and Inglis (2009a) suggested that argument presentation and argument comprehension may be key activities during students' learning activities and students' assessment at university and that there is little knowledge regarding both. Based on the findings of the review, research on both situations remains scarce.

Analyzing the connections between the different aspects, it becomes clear that research focuses largely on specific combinations of resources, processes, and situations. Thus, combinations and interactions of the different perspectives are seldom studied in depth; a limitation regarding the breadth of research that could already be observed regarding the individual aspects. Furthermore, in comparison to other domains, mathematical argumentation and proof skills appear to be rarely researched in a comprehensive way incorporating multiple processes, resources, or situations (see Erduran & Jiménez-Aleixandre, 2008; Khine, 2012), possibly due to a large part of research more directed towards proof as a concept than proof as a process or individual skill (see Reid & Knipping, 2010). Still, Mariotti (2006, p. 173) sees a shift in research, expecting future research to focus less on the conceptions of proof.

The employed bubble charts (see Figures 36-38) offer a clear view of many connections between the three aspects that were not represented in the included research reports. Obviously, not all combinations are of the same importance, but from a theoretical view several further connections should be interesting. For example, the relation of the resource *problem-solving skills* to *problem identification* and *questioning* as initial processes during problem solving, conjecturing, and proof construction, might be an interesting connection, as these processes may shape the whole process.

Of course, the results of this review regarding the foci within the three aspects and their combinations must be handled with care: First, although partially backed up by similar findings in other reviews (Matos, 2013; Mejía-Ramos & Inglis, 2009a), the results of a review are certainly somewhat dependent on the inclusion and exclusion of different sources for papers. Focusing on the proceedings of the International Group for Psychology of Mathematics Education bears the danger of overlooking research that is published elsewhere. Yet, this is also the case when journals are used as sources.

Second, the gaps identified in the reviewed literature, especially within the combinations among the different aspects, can have several reasons. One possible reason, although unlikely in our eyes, is that several (sub-)aspects of mathematical argumentation and proof skills are already well understood so that they are no longer in the focus of research. Another possible reason is publication bias (e.g., Kühberger et al., 2014), leading to fewer publications on certain aspects, respectively combinations of aspects, which show only marginal effects. For example, it may well be that researchers have addressed the influence of mathematical strategic knowledge on proof reading and it turned out to be only marginal (cf. Hodds et al., 2014) and therefore no such papers were published.

Third, the frameworks used to structure the three aspects (resources, processes, and situations) have to be seen as "in development". Although the frameworks are widely accepted, there is few research on their adequacy and categorizations using other aspects or dimensions may lead to somewhat different results.

Overall, our literature review reveals that with 20% of all included research reports, there is a sound basis of research on mathematical argumentation and proof. Still, the perspectives taken on mathematical argumentation and proof skills in the reviewed research reports are often restricted to specific aspects such as single epistemic activities or one or few resources and mostly concentrate on argument and evidence construction. However, multiple theoretical accounts describing argumentation and proof processes within and outside of mathematics (e.g., Boero, 1999; Fischer, Kollar, et al., 2014; Mejía-Ramos & Inglis, 2009a; Schoenfeld, 1985) emphasize that handling mathematical argumentation and proof comprises more processes, is needed in more situations, and possibly requires more resources than explored so far. Here, theoretical models connecting the three aspects *resources*, *processes*, and *situations* are needed, describing for example the role of the resources in different processes and which processes are employed in which situations. Furthermore, the findings of this descriptive review reveal several gaps within recent research. These should be reflected on a theoretical basis regarding their importance for argumentation and proof processes and then analyzed empirically, also appreciating well-designed studies that show negative, no, or minimal effects. Overall, it may be time to build on the existing basis of research on the resources, processes, and situations and to start studying the relations and interactions between the different facets of mathematical argumentation and proof skills to obtain a coherent picture as well as more detailed knowledge on approaches to foster students' mathematical argumentation and proof skills effectively.

The findings from the review further suggest that there is mostly qualitative research on argumentation and proof skills in mathematics education, a finding that reflects a more general predominance of qualitative methods in mathematics education research (Matos, 2013), and that the percentage of qualitative research is especially high when taking broader perspectives on mathematical argumentation and proof skills. Thus, up to now research on mathematical argumentation and proof skills is mostly qualitative in nature, enabling the generation of theoretical models as well as hypotheses, for example regarding the influence of the resources.

Although this research is needed in initial steps, small-scale studies are often not reasonable to create further reliable scientific evidence for or against these hypotheses or models, nor can they be used to measure the impact of the individual resources on students' performance. They therefore only give initial evidence that should be complemented by future quantitative research that allows examining the individual influences of resources or their interactions in more detail. Corresponding quantitative results can help to ascertain or, where needed, adjust findings from theoretical and qualitative research and vice versa, theoretical and qualitative research can build on these results and interpret and integrate them into prior models and conceptions. And although taking a broader perspective of mathematical argumentation and proof skills in quantitative research will pose methodological problems, for example regarding sample size or time for testing, it is important to find ways to study the complex interactions of the often disconnected aspects described in existing research.

Another aspect currently not addressed by research is the impact of *control* or *self-regulation* (e.g., De Corte et al., 2000; Schoenfeld, 1985, 1987) on mathematical argumentation and proof processes and how these can be integrated in models for either their resources or also their processes, for example as *meta-process*es, controlling, selecting, and mediating between the other processes. So far, the coordination of various processes during mathematical argumentation and proof activities is mainly unclear.

Concluding, the relative importance of different situations and different epistemic activities in mathematics education research has to be seen in conjunction with the overall goals of mathematics instruction. These were and still are more focused on argument construction. Still, we cannot expect to gather a comprehensive understanding of mathematical argumentation and proof skills while having blind spots in our research. A broader view on mathematical argumentation and proof skills incorporating resources, processes, and situations can also help to reconsider if the exclusive focus on argument construction is a desirable and viable goal for mathematics education.

5.2 The Impact of Individual Cognitive Resources on Students' Mathematical Argumentation and Proof Skills

5.2.1 Introduction

Argumentation[21] skills represent an increasingly important learning goal across domains (e.g., Berland & McNeill, 2010; Halpern, 1998; National Research Council, 2012) and are often labeled as 21st century skills (e.g., Trilling & Fadel, 2009). In mathematics, domain-specific argumentation skills have been an important educational goal for decades (Hanna & Jahnke, 1996) and are part of standard documents and curricula for secondary education worldwide (Common Core State Standards Initiative, 2010; KMK, 2012). At university level, these skills become even more crucial and represent one of the most important learning goals (Alibert & Thomas, 1991; Jones, 2000; National Council of Teachers of Mathematics, 2000), because here mathematical proofs are introduced as the core method of mathematics as a scientific discipline. Research has repeatedly revealed that not only secondary school students (e.g., Healy & Hoyles, 2000), but also university students have severe problems with handling proof, be it with the construction of proofs (Weber, 2003) or the validation of given purported proofs (A. Selden & Selden, 2003).

One reason for these difficulties and the large variation in students' performance in handling proofs can be related to the limited availability of multiple cognitive resources underlying their performance in constructing and validating proofs (Blömeke et al., 2015). Although prior research has suggested a number of such resources (e.g., Schoenfeld, 1985; Weber, 2001), for example conceptual mathematical knowledge, problem-solving skills, metacognition, or mathematic strategic knowledge, their relative importance for a learner's mathematical argumentation and proof skills remains mainly unclear. Some of these resources and their relative importance on overall mathematical argumentation and proof skills have been studied intensively in the past (e.g., conceptual knowledge; Chinnappan, Ekanayake & Brown, 2012), while others (e.g., mathematical strategic knowledge; Weber, 2001) have not yet been thoroughly addressed.

Starting from a theoretical description of the cognitive resources necessary for handling proofs and the way they influence students' performance, we present an empirical study that sheds light on the relative influence of various individual cognitive resources on students' performance in handling proofs, more specifically in the construction of proofs and the validation of given, purported proofs. We highlight commonalities and differences between both, proof construction and validation, and specifically evaluate the influence of domain-specific and domain-general resources on mathematical argumentation and proof skills. Our main goal is to identify resources that are needed to show high performance in handling mathematical proof tasks. Besides giving theoretical insights this may also be the fundament to create effective means for supporting students' acquisition of mathematical argumentation and proof skills.

5.2.2 Conceptualizing Mathematical Argumentation and Proof Skills

5.2.2.1 Argumentation vs. Proof

As mathematics is a proving science (Heintz, 2000), argumentation and proof are ubiquitous in mathematics and learning to deal with proofs constitutes an important research area in

[21] Parts of this study and preliminary analyses have been presented at the conferences PME 40 and GDM 2016 and published in the respective proceedings volumes (Sommerhoff, Ufer, & Kollar, 2016b, 2016c).

mathematics education. Still, there is no universally accepted definition of mathematical reasoning, argumentation, and proof and often enough, they are not clearly defined (Reid & Knipping, 2010). In this article, we define argumentation analogous to Halpern (2002) and Toulmin (2003), as a sequence of (in general not necessarily deductive) inferences, leading from given premises to a conclusion, aimed at providing evidence for or against a given claim (Reichersdorfer et al., 2012; Reiss et al., 2008) or exploring a given task, problem, or situation (e.g., Reiss & Ufer, 2009).

Building on this, a proof can be defined as a mathematical argumentation that is subject to several characteristics and restrictions (A. J. Stylianides, 2007), often referred to as socio-mathematical norms (Yackel & Cobb, 1996). Stylianides (2007) names three categories of such characteristics that is a *set of accepted statements*, *modes of argumentation,* and *modes of argument representation* (see further methodological knowledge by Heinze and Reiss (2003) for an alternative conceptualization). Although socio-mathematical norms are specific to a social context, norms concerning proof are considered to be highly consistent internationally (see Dawkins & Weber, 2016; Heintz, 2000): Only deductive arguments are permitted (*modes of argumentation*). These deductions have to build on an underlying, axiomatic framework and to use only those statements as warrants that were already deduced and are accepted as "known" within the social context (*set of accepted statements*). Furthermore, the arguments of a proof have to be written out in such detail that the proof is in principle formalizable (Alama & Kahle, 2013). Yet, this degree of formality varies widely, especially from early to more advanced lectures.

5.2.2.2 Proving as Problem Solving

Proving, that is the process of constructing mathematical proofs, is often conceptualized as a problem-solving process (e.g., Polya, 1945) since students normally do not have direct means to construct a proof and this inaccessibility of a direct solution is often used to define a problem (e.g., Dörner, 1979; Mayer, 1983; Schoenfeld, 1985). The process of proving itself is then often regarded from an information-processing perspective (J. R. Anderson, 1993; Simon, 1978), characterizing it as a (dual) search space (Klahr & Dunbar, 1988; Newell & Simon, 1972). In short, problem solvers need to create an appropriate representation of the problem (problem space) and use available operations (e.g., experimentation, restructuring, algebraic computations, deductions) to navigate within that representation to progress from the starting point to the solution of the problem. Here, special emphasis lies on the creation and structuring of the problem space(s) and the subsequent identification and selection of appropriate operations that help the problem solver to get closer to the solution.

5.2.2.3 Mathematical Argumentation and proof skills as a Latent Variable

Building on these definitions and conceptions, the present study analyzes students' mathematical argumentation and proof skills. Following theoretical considerations (Blömeke et al., 2015) we consider students' mathematical argumentation and proof skills as a latent variable, which becomes visible in students' individual performance in different, specific situations that involve dealing with mathematical argumentation and proof. This performance is observable, can be judged against local, socio-mathematical norms and can serve as an indicator for their mathematical argumentation and proof skills (see Figure 39, right side). It comes along through the application of students' mathematical argumentation and proof skills during a process of coping with the demands of the situation. How well students can cope with these demands is largely determined by the availability of cognitive resources, such as domain-specific knowledge or problem-solving skills.

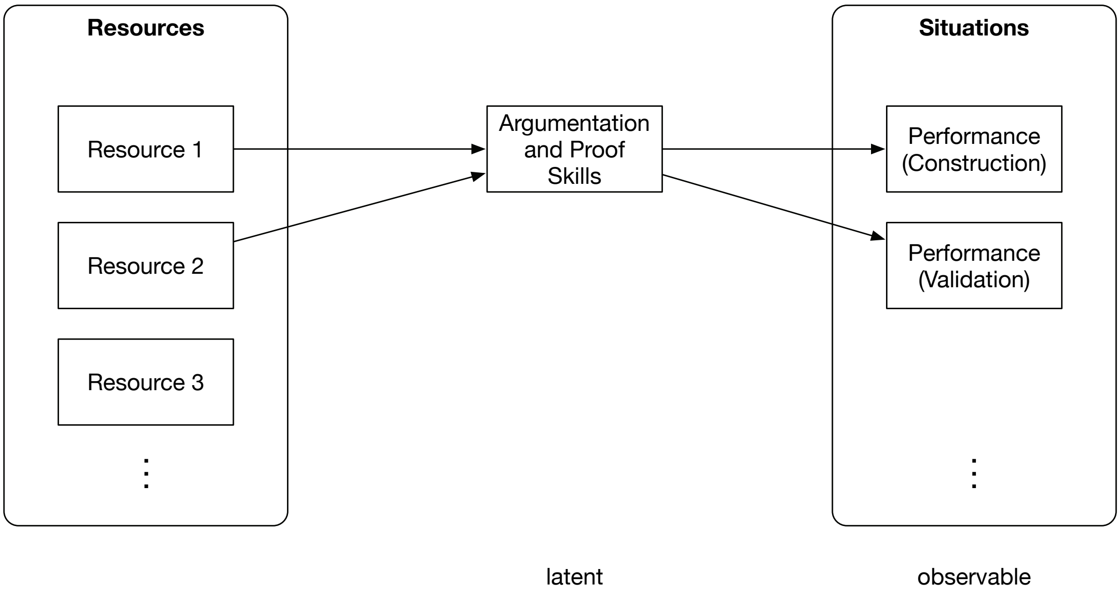


Figure 39. Relation between students' resources, their mathematical argumentation and proof skills, and the situations posing different demands on students (based on Blömeke et al., 2015).

Accordingly, students' cognitive resources can be used to explain their performance in different situations and therefore allow inferences on their mathematical argumentation and proof skills, which cannot be measured directly. The resulting data gives insights on how successful students differ from less successful students, and which resources are most important when dealing with argumentation and proof in various ways. Thus, the information is vital as a basis to understand students' achievements and problems in handling proof.

5.2.3 Situations Related to Mathematical Argumentation and Proof

When working with proofs, students can be faced with different situations posing different demands: Students may be asked to construct a proof or to think through an already existing proof with the aim of understanding the underlying arguments, for example to answer questions about the proof or to provide feedback to a peer. These various demands occur in the form of learning or assessment tasks, or during their mathematical work, for example when trying to find and argue for solutions to problems within or outside mathematics. Up to now, research in mathematics education mostly focusses on the situation of proof construction, yet other situations come to the fore slowly (e.g., Alcock & Weber, 2005; Inglis & Alcock, 2012; A. Selden & Selden, 2003, 2015a; Weber, 2008).

Mejía-Ramos and Inglis (2009a, 2009b) proposed a framework based on the work of Giaquinto (2005) that highlights three main types of situations in the context of mathematical argumentation and proof: The *construction of a novel proof*, the *reading of a given proof,* and the *presentation of a given proof*.

For proof reading, the subcategories by Mejia-Ramos and Weber were later taken up by A. Selden and Selden (2015a) in a qualitative research review, coining the terms *proof comprehension*, *proof validation,* and *proof evaluation* for these three subcategories. Here, proof comprehension refers to the activity of reading a proof that is known to be true (e.g., from a reliable textbook) with the aim of understanding it. Proof validation and proof evaluation, in contrast, refer to the reading of a (purported) proof from a questionable source. This could, for example, be a proof constructed by a fellow student. The goal of proof validation is to judge a proofs' correctness, relative to the current, local, socio-mathematical norms, whereas proof evaluation also includes a judgement regarding further criteria such as beauty or simplicity.

Although the three main categories – construction, reading, and presenting – are per se equally important, proof construction and proof reading dominate learning and assessment in most university programs as well as research in mathematics education (Mejía-Ramos & Inglis, 2009a). For example, students spend a major amount of time on their exercise sheets or in tutorials accompanying their lectures, both mostly focusing on proof construction. Besides construction, students also spend considerable time on reading proofs, here mostly on proof comprehension, for example when learning from books, or proof validation, for example when reading a peer's purported proofs for an exercise sheet.

This focus on proof construction and reading is reflected in research on mathematical argumentation and proof. In their literature review of 131 quantitative research articles, Mejía-Ramos and Inglis (2009a) found 82 papers (73 %) addressing proof construction, 24 papers (21%) addressing proof reading and zero papers (0%) addressing proof presentation. Our review on 782 research reports of the proceedings of the Annual Conferences of the International Group for the Psychology of Mathematics Education (see section 5.1) revealed similar numbers.

In summary, students are faced with different situations that require their mathematical argumentation and proof skills and therefore also their underlying resources. Up to now, it is unclear if and how the usage of the resources differs between these situations.

5.2.4 Resources Underlying Mathematical Argumentation and Proof Skills

In the context of constructing mathematical argumentations and proofs, several frameworks, as well as studies, document the influence of certain resources like domain-specific knowledge or problem-solving skills on students' performance (Chinnappan et al., 2012; Hellmich, Hartmann, & Reiss, 2002). However, similar research on the impact of these and other resources on proof validation and proof evaluation performance is scarce (e.g., Inglis & Alcock, 2012; Ko & Knuth, 2013) and only vaguely mentions resources such as domain-specific content knowledge or specific mathematics-related strategies. Thus, research on the resources underlying mathematical argumentation and proof skills so far does not distinguish between different situations although these may reasonably require different resources for a successful handling of proof.

Reviewing general frameworks for problem solving (e.g., Schoenfeld, 1985) or self-regulated learning (e.g., De Corte et al., 2000), mostly cognitive and affective resources have been discussed, often accompanied by beliefs (Op 't Eynde et al., 2002). Yet, although affective aspects and beliefs are repeatedly highlighted as crucial for argumentation and proof construction processes (e.g., Furinghetti & Morselli, 2009; Schoenfeld, 1983), empirical studies so far were not able to show consistent results (e.g., Heinze & Reiss, 2009). This may result from ambiguities in the definition of the diverse affective constructs (Furinghetti & Pehkonen, 2002; Grootenboer & Marshman, 2016; Hannula, 2006) or methodological issues regarding especially the measurement of state characteristics (Hannula, 2006; Törner, 2002). Further, multiple studies gave evidence for the high stability of students' beliefs and attitudes towards mathematics (e.g., Kloosterman et al., 1996).

In contrast, several studies (e.g., Chinnappan et al., 2012; Ufer et al., 2008) were able to show the high relevance of cognitive resources such as specific knowledge facets or problem-solving skills. In consequence, the present study focuses solely on cognitive resources in order to create a more nuanced picture.

A number of cognitive resources underlying students' mathematical argumentation and proof skills are mentioned in existing literature. Among those, six resources stand out, since they either appear in multiple frameworks or their influence has been established on strong theoretical or empirical grounds.
In analogy to Chinnappan and Lawson (1996) these resources can be broadly categorized as *content-specific*, *domain-specific*, and *domain-general* resources. Here, *content-specific* refers to resources that are bound to a specific content area and are therefore heavily restricted in their range of applicability. An example is the definition of a Cauchy series within the content area of infinite sequences. In contrast, *domain-specific* resources are not tied to a specific content area, but allow for near transfer to other content areas, problems, or situations within mathematics. An example is the Gaussian elimination algorithm. Finally, *domain-general* resources are characterized by their wide applicability throughout different domains. That is, they allow for transfer to a broad set of situations in domains such as mathematics, biology or history. An example is the strategy of means-end analysis.
So far, most empirical studies did not study these resources in combination, so that little is known about their interplay and their relative importance. Furthermore, up to now, no comparisons between their influences within different situations involving mathematical argumentation and proof skills were made.

5.2.4.1 Conceptual and Procedural Mathematical Knowledge

Two fundamental content-specific resources discussed as underlying students' mathematical argumentation and proof skills (Ufer et al., 2008) are conceptual and procedural mathematical knowledge (Hiebert & Lefevre, 1986). The necessity of underlying mathematical knowledge can be seen as in line with several developmental theories like Piaget's stages of cognitive development (Inhelder & Piaget, 1958) or van Hiele's stage theory of understanding in geometry (Fuys, Geddes, & Welchman Tischler, 1984; van Hiele & van Hiele-Geldorf, 1978), which regard stages of increasing depth of conceptual knowledge as prerequisites for the development of more complex skills, such as mathematical argumentation and proof skills, at higher stages. Furthermore, Anderson's theory of cognitive skill acquisition (1982) underlines the importance of specific knowledge for a cognitive skill such as handling proofs. From an information-processing perspective (J. R. Anderson, 1993; Simon, 1978), theorems and definitions constitute problem-solving operators (see Klahr & Dunbar, 1988) that need to be purposefully combined to generate a deductive chain of arguments in proof construction, respectively be matched with a given chain of arguments to comprehend or validate proofs. A more elaborate cognitive representation of these operators as well as relations between them are considered central to direct search processes in the problem space (Chinnappan & Lawson, 1996). Furthermore, rich conceptual knowledge of the concepts involved in a proof, for example, prototypical examples, special cases, non-examples to a given concept, or counter-examples to wrong propositions support semantic evaluation processes while constructing or validating proofs (Lockwood et al., 2016). Beyond these conceptual aspects, procedural knowledge (de Jong & Ferguson-Hessler, 1996; Rittle-Johnson et al., 2015) may serve as a problem-solving operator to restructure the problem representation using routine, technical skills such as algebraic manipulations or calculation techniques. Accordingly, models of self-regulated learning and problem solving (De Corte et al., 2000; Schoenfeld, 1985) emphasize the role of mathematical knowledge to perform complex, domain-specific skills.
So far, several empirical studies provided evidence of the influence of mathematical knowledge on students' mathematical argumentation and proof skills: For example, the studies of

Chinnappan et al. (2012), as well as Ufer et al. (2008), found geometry content knowledge to be the most important predictor of secondary students' geometry proof skills.
In this study, we conceptualize conceptual mathematical knowledge to relate to knowledge about the mathematical content area used in the statement to be proved or in the proof itself. Similarly, procedural knowledge refers to knowledge of technical, routine aspects that are applied in the proof. Examples are the definition of a Cauchy sequence, linked to several other facts regarding infinite series, for conceptual knowledge, and applying the geometric series while solving an equation for procedural knowledge.

5.2.4.2 Mathematical Strategic Knowledge (MSK)

The construct mathematical strategic knowledge (MSK) was first introduced by Weber (2001). In a qualitative study with undergraduate and doctoral students, he observed that although students had the knowledge about concepts and procedures necessary to perform certain proofs, they lacked the mathematical strategic knowledge to apply it appropriately. Reiss and Heinze (2004) found similar results for pupils trying to solve geometry tasks. To address this lack of knowledge, Weber defined mathematical strategic knowledge as "knowledge of how to choose which facts and theorems to apply" (Weber, 2001, p. 101) and highlighted different aspects of strategic knowledge (e.g., knowledge of which theorems are important and when they will be useful, knowledge of when and when not to use 'syntactic' strategies). Weber thus characterizes mathematical strategic knowledge as a domain-specific version of general problem-solving strategies, respectively heuristics. In particular, Weber conceptualized mathematical proof tasks as complex problem-solving tasks (Weber, 2001, p. 111) and justifies mathematical strategic knowledge by using analogies to heuristics in general problem solving (e.g., Abel, 2003). The concept of mathematical strategic knowledge is also related to the more general notion of conditional knowledge (Lehmann & Magidor, 1992), often defined as "knowing when and why to use" certain pieces of knowledge, and is also examined more generally as strategic knowledge in other disciplines (Paris, Lipson, & Wixson, 1983; van Gog, Paas, & van Merriënboer, 2004).
Despite these relations, mathematical strategic knowledge has to be clearly distinguished from these concepts, as it entails only domain-specific strategic knowledge. For example, the strategy "If a proof asks to show an equality for all natural numbers, complete induction could be promising" is neither part of conceptual or procedural knowledge, as it is a heuristic strategy that only allows transfer over several content areas in mathematics, but cannot be classified as a problem-solving heuristic as it only relates to mathematics.
Mathematical strategic knowledge is relevant throughout constructing or validating a proof, each time helping to decide what step to do next or to understand why certain steps are done and judge whether these can be effective. Adopting an information-processing view, mathematical strategic knowledge refers to knowledge about possible operators, which can be used in the problem space of a specific proof, and knowledge about when to apply which of these. It can be seen as especially important immediately after the creation of the problem space, which greatly relies on content knowledge, as it governs the direction of the problem-solving process and will constrain, hopefully in a positive way, the number of possible operations within that problem space.
In this study, we conceptualize mathematical strategic knowledge to relate to knowledge about domain-specific strategies linking certain cues and hints in tasks as well as proofs to mathematical concepts or methods that can be especially useful to tackle the task. For example, using an $\varepsilon/2$-approach when dealing with ε-δ-continuity.

5.2.4.3 Methodological Knowledge (MeK)

Methodological knowledge (Heinze & Reiss, 2003) is suggested as a further resource required for handling mathematical proofs as it entails knowledge about different types of proofs, their nature and purposes, as well as acceptance criteria for mathematical proofs (de Villiers, 1990; Hanna & Jahnke, 1996). Methodological knowledge is dependent on the local mathematical community and contains the standards (socio-mathematical norms) set within the community (see Dawkins & Weber, 2016; Fallis, 2003; Heinze, 2010).
Heinze and Reiss (2003) name three core aspects of methodological knowledge: The *proof scheme*, the *proof structure,* and the *chain of conclusions* relating to the accepted types of inferences, the overall structure of the proof, and the logical sequence of neighboring inferences and arguments.
Up to now, the importance of methodological knowledge was emphasized by several studies (e.g., Healy & Hoyles, 2000; Heinze & Reiss, 2003; Ufer, Heinze, Kuntze, et al., 2009), yet their assessment is unsatisfactory: So far, students' methodological knowledge was assessed indirectly using their performance in proof validation tasks, which can be assumed to be dependent on methodological knowledge, yet is likely to also have other underlying resources and therefore only partially reflects methodological knowledge (see Figure 39).

5.2.4.4 Problem-solving Skills (PSS)

General problem-solving skills are widely discussed in the context of proof (Chinnappan et al., 2012; A. Selden & Selden, 2013; Ufer et al., 2008; Weber, 2005), as constructing proofs is often conceptualized as a problem-solving process. Here, the availability of domain-general problem-solving heuristics (Polya, 1945) and the ability to apply these purposefully, as well as self-regulation and metacognition are in focus. The influence of heuristics and the ability to apply them has been stated as an important factor in several theoretical frameworks (e.g., Schoenfeld, 1985) and also been empirically underpinned by a few studies in secondary school geometry classrooms (e.g., Chinnappan et al., 2012; Ufer et al., 2008). Although there are doubts about the transfer of these strategies, their domain-generality, and therefore their broad effectiveness (Sweller, 1990; Tricot & Sweller, 2014), numerous studies underlie their benefits (e.g., Chinnappan & Lawson, 1996; D. W. Eccles & Feltovich, 2008).
In this study, we conceptualize problem-solving skills as the knowledge about domain-general heuristics as well as the ability to apply them in content-independent situations such as the three jug problem.

5.2.4.5 Metacognitive Awareness (MA)

Metacognition is another repeatedly mentioned resource in the context of mathematical argumentation and proof skills (De Corte et al., 2000; Reiss et al., 2002; Schoenfeld, 1985, 1987; Zohar & Peled, 2008). Metacognition refers to knowledge and cognition about cognition as introduced by Flavell (1979) or put differently, to the monitoring and regulation of cognition. Most models of problem solving and self-regulation assign a central role to metacognition for learning of and performance on complex tasks (De Corte et al., 2000; Griffin, Wiley, & Salas, 2013), including proof construction (Heinze & Reiss, 2007). Kramarski et al. (2002) as well as Lingel, Neuenhaus, Artelt, and Schneider (2014) were able to show a positive influence of metacognition on students' performance in mathematical tasks and also Lockwood et al. (2016) mention its importance when exploring and proving conjectures. Although metacognition plays an important role in problem solving, is regarded as domain-general and is sometimes seen as a specific part of problem solving, it cannot be seen as equal to problem-solving skills or heuristics, respectively, as it is, by definition, not restricted to the context of problems. Further,

metacognition is assumed to be domain-general and concerned with monitoring and knowledge of cognition in general, thereby unlike mathematical strategic knowledge that is specific to mathematics. Yet, the latter might be utilized in metacognitive activities. In the context of proof, metacognition should be especially important when checking if the current step during the construction or validation of a proof aligns with the preceding steps and whether the argumentation is still directed towards the desired conclusion.
In this study, we conceptualize metacognitive awareness as the degree to which students are aware of their ability to reflect upon, understand, and control their learning.

5.2.4.6 Conditional Reasoning Skills (CRS)

Probably the most central norm for the acceptance of an argumentation as a proof is the sole use of deductive inferences within the argumentation. Accordingly, conditional reasoning skills, that is being able to correctly make or judge different types of deductive inferences, for example modus ponens or tollens ("If it is raining, the sky is cloudy. The sky is not cloudy, therefore it cannot be raining"), can be hypothesized to exert a positive influence on an individual's mathematical argumentation and proof skills (Heinze & Kwak, 2002). Several studies have researched students' skills in conditional reasoning and the effects of explicit teaching (e.g., Inglis & Simpson, 2008, 2009), often motivated by the theory of formal discipline (Attridge & Inglis, 2013). Some qualitative studies have underlined that this kind of formal reasoning is actually used in proof construction and proof validation (Weber, 2008). Yet, neither an empirical relation of conditional reasoning skills to proof construction, nor to proof validation skills has been explicitly shown so far.
In this study, we conceptualize conditional reasoning skills to relate to the skill to correctly accept or reject logical inferences such as modus ponens, modus tollens, affirming the consequent, and denying the antecedent.

Summarizing, all presented cognitive resources are regarded as important resources for mathematical argumentation and proof skills and were already in part empirically established as such. However, all of these resources are likely to be positively correlated and therefore the existing evidence on their influence on mathematical argumentation and proof skills is likely to be flawed. Up to now, there are very few studies that examine several of these resources simultaneously (see section 5.1). Accordingly, it is unknown, whether the various potential resources can be empirically separated and what their relative influence on the performance in different proof related situations such as proof construction and validation is.

5.2.5 Connection Between Proof Validation and Proof Construction

Although proof construction and proof validation are clearly different activities, with a purported proof being the *prerequisite* of proof validation and the *goal* of proof construction, there are several links connecting both activities (e.g., Pfeiffer, 2009b, 2011; Powers, Craviotto, & Grassl, 2010; A. Selden & Selden, 2003), building mainly on two theoretical arguments.
First, taking the perspective of resources underlying proof construction and validation (Figure 39), a relation between both can be expected as some resources are necessary to cope with both kinds of situations. For example, conceptual mathematical knowledge should be necessary for both activities: for proof construction to give a warrant for a certain inference and for proof validation to understand a certain inference and accept it based on its warrant. Also, mathematical strategic knowledge should be valuable for both activities. It guides proof construction in the hopefully correct direction and informs the validation process that a certain approach taken in the purported proof might not be optimal for the given task. Still, both cannot

be expected to completely draw on the same resources and especially not to the same extent. For example, problem-solving skills may be more relevant for constructing a proof than for validating a given proof. Further, during proof construction students need to identify and possibly connect the relevant conceptual knowledge themselves, whereas during proof validation these are pre-given and just need to be checked for their correctness and correct implementation.
Second, both construction and validation can be linked on an activity level. Process models of problem solving (Polya, 1945), scientific reasoning (Fischer, Kollar, et al., 2014), and also constructing mathematical proofs (Boero, 1999) give a more or less explicit reference to validation as a part of these processes. Polya mentions it as “looking back”, Fischer et al. include it as “evidence evaluation”, and also Boero mentions in his expert model for proof construction that the reliability of arguments has to be checked. Thus, all three models include the validation of the prior steps taken while solving a problem, creating a coherent scientific argumentation or constructing a mathematical proof as a kind of monitoring activity. Accordingly, proof validation can be rightfully seen as an important sub-process of proof construction.
Empirical findings support these arguments, showing positive correlations between students’ performance in proof construction and proof validation. In a study on geometry proof skills with almost 700 students in 8^{th} grade Ufer, Heinze, Kuntze, et al. (2009) found that students’ validation skills explained 10% of the variance in students’ proof scores, after controlling for performance on geometry calculation tasks. Thus, both proof construction and proof validation are partially connected, but the low explained variance also emphasizes that both also differ vastly.
The correlation between both performances can be interpreted in several ways. Proof validation tasks have been put forward as measures of students’ *methodological knowledge*. Accordingly, the observed correlation between proof validation and proof construction regarding the same mathematical concepts may be due to several factors, such as content-specific knowledge or more general knowledge about acceptance criteria. It remains an open question to which extent there is a relation between validating proofs in one content area and constructing proofs in another content area. A positive relation, when controlling for conceptual knowledge, may be a first indication that students indeed transfer knowledge about acceptance criteria from one content area to another and methodological knowledge underlying proof construction and proof validation is responsible for their correlation.

5.2.6 The Current Study

Several studies, for example by Healy and Hoyles (2000) or A. Selden and Selden (2003) reveal not only that students have problems in proof construction as well as proof validation throughout different age groups, but also show substantial interindividual variation in students’ performance on proof-related tasks. Summarizing the theoretical model described above, mathematical argumentation and proof skills are assumed to base on certain cognitive resources students have acquired during their education. Prior evidence indicates that these resources strongly influence a students’ chance to be successful when asked to construct, validate or comprehend a given proof.
The main goal of the current study was to explore the relative influence of individual domain-specific and domain-general resources on students’ proof construction and validation performance. Based on prior research, we selected six different cognitive resources, which have been discussed as important prerequisites of students’ mathematical argumentation and proof skills: Conceptual mathematical knowledge, procedural mathematical knowledge, mathematical strategic knowledge, problem-solving skill, metacognitive awareness, and conditional reasoning skills. Here, we deliberately excluded methodological knowledge as current assessment relies on

students' proof validation performance as a proxy, which will be examined more generally within this study.

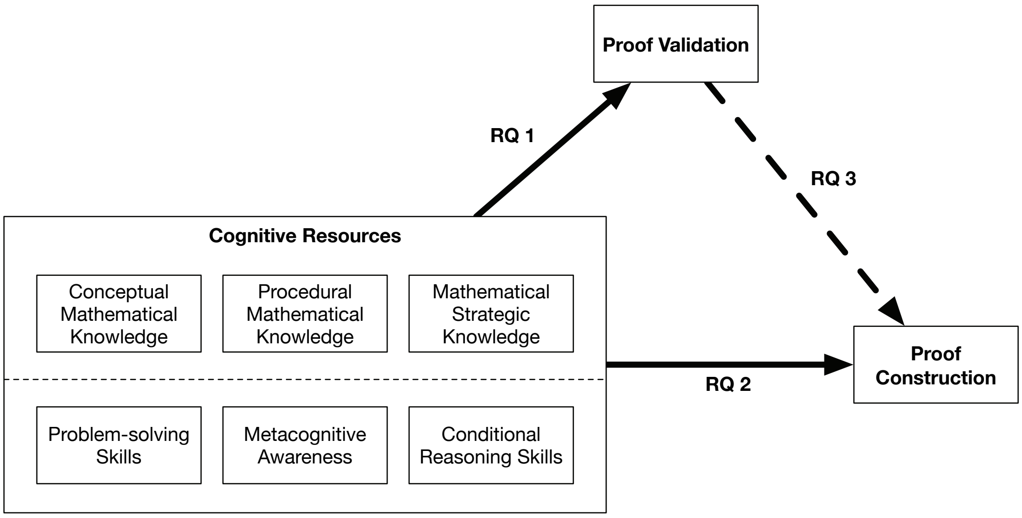


Figure 40. Relations between resources and proof validation (RQ1), resources and proof construction (RQ2), and proof validation and proof construction (RQ3).

First, we were interested in the role of these resources regarding proof validation and proof construction (solid lines in Figure 40).

RQ1 What is the relation of the six cognitive prerequisites (conceptual and procedural mathematical knowledge, mathematical strategic knowledge, problem-solving skills, metacognitive awareness, and conditional reasoning skills) with students' performance in *proof validation*, when controlling for the other resources?

Based on the role of the selected resources described above and in prior research, we expected a non-negative relation of all resources to proof validation performance. Specifically, we expected pronounced relations of basic conceptual mathematical knowledge and conditional reasoning skills, as we expected students to check the individual facts and inferences within the given purported proofs.

RQ2 What is the relation of the six cognitive prerequisites with students' performance in *proof construction*, when controlling for the other resources?

Following literature and prior studies, especially prior regression analyses by Chinnappan et al. (2012) and Ufer et al. (2008), we expected a non-negative influence of all resources, specifically for conceptual knowledge and problem-solving skills. Further, we expected an especially strong relation of procedural mathematical knowledge with proof construction, since proof construction primarily requires students to perform technical skills themselves and structure the problem underlying the proof tasks. We also expected that the influence of conditional reasoning skills would be higher as proof construction can be interpreted as a complex search process for a deductive chain of arguments.

Furthermore, theoretical arguments suggest a relation between students' performance in proof validation and in proof construction. Still, so far it is unclear whether this is due to content-specific resources underlying both activities, or whether there is a content area overarching independent resource, such as knowledge about what constitutes an acceptable mathematical proof (dashed line in Figure 40).

RQ3 To what extent can students' proof validation performance explain variance in students' proof construction performance over and above the six resources?

As both, proof validation and proof construction performance are cognitive performance measures relying on students' mathematical argumentation and proof skills, a positive correlation can be expected and has already been shown in prior research (Ufer, Heinze, Kuntze, et al., 2009). Beyond that, given that proof validation performance is assumed to depend on content-overarching methodological knowledge about acceptance criteria for mathematical proofs, one would expect it to explain additional variance in students' proof construction performance as this knowledge facet is not included in the six resources. This would also reflect process models of proof construction (e.g., Boero, 1999) or scientific reasoning (e.g., Fischer, Kollar, et al., 2014), implying that proof validation is a cognitive sub-process of proof construction that at least partially relies on domain-specific but content-overarching knowledge.

5.2.6.1 Design and Sample

The analysis presented here is part of a larger study with students from a German university, which took place in a voluntary course labeled "Mathematical proofs: That's how to do it!". For this study, we analyzed data from the first measurement in the first session of the course. From the total number of participants two had to be excluded due to sloppy completion of one scale. The final sample consisted of 64 participants (40 female, 23 male, 1NA), which were in their first (49) respectively third semester (14) of either a bachelor's program in mathematics or business mathematics (48) or a mathematics teacher education program for secondary schools (Gymnasium, 14). Two students did not indicate their degree program, one did not indicate the semester. At the time of data collection, all participants had attended a proof-based real analysis lecture, covering limits of number sequences and infinite series, functions, and continuity.
In the first session of the larger study, we asked participants to complete questionnaires including subscales for each cognitive resource and items on proof construction as well as proof validation, respectively. Furthermore, they were asked to fill out additional instruments, which are not within the scope of this paper.

5.2.6.2 Instruments

To measure students' individual cognitive resources as well as their performance in proof construction and proof validation, we adapted several already existing scales and created new scales. This step was necessary because either no suitable scale for the resources existed (e.g., mathematical strategic knowledge) or the scales had to be adapted to the university mathematics context (e.g., metacognitive awareness).
Proof construction: Students' performance in constructing proofs was assessed using four prototypical tasks from a proof-based real analysis lecture. They were chosen spanning several different types of proof (proof by induction, proof by contradiction, direct proof) and were, at least in their specific formulation, unknown to the students, although similar items had been part of their lectures. For each item, students were asked to prove the provided statement and give reasons for their steps within the proof. Their written answers were scored on a multi-level scale (0-4) based on the occurrence of the main steps required for the proof.
Proof validation: Items for proof validation focused on proofs from elementary number theory - a different content area than for proof construction. The items consisted of a proof task as well as four purported proofs from students. Each of the proofs was written in a concise, semi-formal way to avoid an overly use of mathematical symbols or narrative parts. One of the four purported proofs was correct, the other three each contained an error in one facet of methodological

knowledge (Heinze & Reiss, 2003), that is an error in the proof scheme, the proof structure, or the logical chain. Students were asked to read all four purported proofs and subsequently answer three closed questions on features of the proof as well as judge its overall correctness in a closed question. Additionally, they were asked to provide an explanation for their decision on the correctness in an open question. Both answers from the closed and the open item regarding the correctness of the purported proof were processed in a dichotomous consistency rating.

Proposition

The product of three arbitrary consecutive integers is divisible by 3.

Martin's proof
Let $a \in \mathbb{Z}$ be an arbitrary integer. Accordingly the two consecutive integers can be written as $a+1$ and $a+2$.
We are interested in the product

$$a \cdot (a+1) \cdot (a+2)$$

Expanding the term yields:

$$a \cdot (a+1) \cdot (a+2) = a \cdot (a^2+3a+2) = a^3+3a^2+2a$$

Since the proposition should hold for an arbitrary integer a, the statement has to hold independently of a. We therefore only look at the coefficients of the term a^3+3a^2+2a. For the sum of these coefficients we get:

$$1+3+2=6$$

Accordingly since $3|6$ holds, it also holds that $3|a^3+3a^2+2a$ resp. $3|a \cdot (a+1) \cdot (a+2)$. Thus the product of three arbitrary consecutive integers is divisible by 3 and we have proven the proposition.

Figure 41. Example item for proof validation with a purported proof containing an error in the logical chain (translated).

Procedural and conceptual knowledge: Because of the two content areas used in proof validation and construction, it was necessary to assess conceptual and procedural mathematical knowledge in both content areas separately. We therefore employed four scales for mathematical knowledge, which were carefully focused on the concepts used in the corresponding proof tasks: *basic conceptual mathematical knowledge (BCMK)* and *basic procedural mathematical knowledge (BPMK)* from elementary number theory as well as *advanced conceptual mathematical knowledge (ACMK)* and *advanced procedural mathematical knowledge (APMK)* from real analysis. The two advanced mathematical knowledge scales were based on scales by Rach and Heinze (2016) and Wagner (2011). The basic mathematical knowledge scales were based on standardized nation-wide tests for school mathematics (Blum, Drüke-Noe, Hartung, & Köller, 2012).

Problem-solving skills (PSS): Following the theoretical considerations above, we surveyed students' general problem-solving skills with tasks that were knowledge-lean (e.g., three jug problem) thus all participants could be expected to have the required conceptual knowledge at hand. In particular, the employed items did not require mathematical knowledge as measured in the four scales described above, except for basic arithmetic calculations. The difficulty of the items was therefore largely based on the application of domain-general heuristics. These were mostly needed for the generation of an appropriate problem space and the identification of suitable operations.

Mathematical strategic knowledge (MSK): To our knowledge, there was no quantitative instrument to assess mathematical strategic knowledge at the start of our project. We therefore developed a questionnaire consisting of four main items, each holding four sub-items. For each main item, a typical task from a proof-based real analysis lecture was provided as well as four excerpts of the task description. Students were then asked to select those excerpts, which gave the most important cues on how to tackle the task (closed format). Additionally, they were asked to explain briefly (open format) which method or concept was hinted at by this information and how it could be used to solve the task. Items were scored as correct when both parts (closed and open) corresponded and described a meaningful way to approach the given task. That is, the

approach could in principle be used to solve the given tasks, even if it might not have worked in the end due to reasons not foreseeable in the task description.
Finally, the scales for *conditional reasoning skills (CRS)* (Evans et al., 1995; Inglis & Simpson, 2008, 2009) and for *metacognitive awareness (MA)* (Schraw & Dennison, 1994) were taken from literature, translated and adapted to the context.

Prior to the use within the reported study, all scales had been piloted and evaluated for their goodness. The reliability of the scales was $.58 < \alpha < .81$, where the .58 corresponds to the only scale below .6 (mathematical strategic knowledge). As the scale for mathematical strategic knowledge has been newly developed and is the first of its kind, the scale was deemed to be acceptable at this point. Over 15% of the open items were coded by two independent coders, resulting in good interrater reliabilities ($\kappa > .76$; $\kappa_{\text{Mean}} = .93$; $SD = .09$).

5.2.6.3 Statistical Analyses

To analyze the impact of the six resources on students' performance on proof validation (RQ1), we calculated Generalized Linear Mixed Models (GLMMs) (Bolker et al., 2009; Zuur, Ieno, & Saveliev, 2009) using the R packages *LME4* (Bates, Mächler, Bolker, & Walker, 2015). GLMMs have several benefits over classical linear regressions. Primarily, they can handle non-normally distributed dependent variables as for example binary variables (used for proof validation) or ordinal variables (used for proof construction) using logistic link functions similar to Item Response Theory models. Thus, they allow for an analysis of the relation of predictors with performance on single items, without having to use scale mean values. This increases the power to detect relations of the predictors. Furthermore, GLMMs allow for controlling for the dependency of the different observations from a single individual as a random effect. Item difficulty was taken into account as a so-called fixed effects.
The interpretation of regression coefficients in GLMMs is comparable to that in logistic and ordinal regression analyses. Larger coefficients represent a stronger relation of a predictor, and the sign of coefficients represents the direction of the relation. We used z-standardized scores for the resources as predictors. Thus, the coefficients for different predictors in the same model can also be compared meaningfully. Following the approach recommended by literature (e.g., Bolker et al., 2009) we calculated all possible models with the given resources. Based on the AICc (Akaike information criterion with a correction for finite sample sizes) we then selected the optimal model and averaged over those models with a ΔAICc < 4, that is we averaged over those models that were indistinguishable close to the optimal model regarding the AICc criterion.
To analyze the impact of the six resources on students' performance on proof construction (RQ2), we slightly adapted the method as students' performance in proof construction was coded ordinally. Therefore, the according GLMM was calculated using the R package *ordinal* (Christensen, 2015), offering an analysis method analogous to the one used before for the dichotomous proof validation measures.

5.2.7 Results

5.2.7.1 Preliminary Analyses

Descriptive statistics of the scales show an acceptable coverage of the possible range, indicating neither ceiling nor floor effects (see Table 4). The mean difficulty is in a medium range for most scales, only the scale for advanced conceptual mathematical knowledge shows a high difficulty with a mean score of .31.

Table 4. Descriptive statistics of the used scales.

	Number of Items	**Mean**[22]	**SD**	**Min**	**Max**
Basic Conceptual Mathematical Knowledge	11	.43	.20	0	.82
Basic Procedural Mathematical Knowledge	7	.56	.18	0	.94
Advanced Conceptual Mathematical Knowledge	8	.31	.19	0	.75
Advanced Procedural Mathematical Knowledge	5	.40	.18	.09	.94
Mathematical Strategic Knowledge	4	.36	.17	.06	.81
Problem-solving Skills	4	.37	.21	0	.86
Conditional Reasoning Skills	16	.64	.19	0	1
Metacognitive Awareness	52	2.84	0.34	1.90	3.86
Proof Validation	4	.45	.18	.10	.85
Proof Construction	4	.35	.18	0	.93

Intercorrelations between the individual scales (see Table 5) show weak to moderate Pearson correlations smaller than r = .45 between the resource scales. Thus, no severe issues with multicollinearity have to be expected in the following regression analyses.

[22] The scale for metacognitive awareness ranges from 1 to 4, based on the Likert scale type items used to measure it, and mean values of the items are given. All other scales are reported on a scale from 0 to 1 and solution rates are reported.

Table 5. Intercorrelations for the cognitive resources.

	Cognitive Resources						
	BPMK	ACMK	APMK	MSK	PSS	CRS	MA
Basic Conceptual Mathematical Knowledge	.317*	.448***	.292*	.274*	.183	.303*	-.075
Basic Procedural Mathematical Knowledge	--	.282*	.394**	.050	.021	.157	-.031
Advanced Conceptual Mathematical Knowledge	--	--	.433***	.219	.288*	.161	.136
Advanced Procedural Mathematical Knowledge	--	--	--	.077	-.007	.284*	.183
Mathematical Strategic Knowledge	--	--	--	--	-.108	-.085	.013
Problem-solving Skills	--	--	--	--	--	.295*	-.085
Conditional Reasoning Skills	--	--	--	--	--	--	-.005

$^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$

5.2.7.2 The Influence of the Resources on Proof Validation Performance

Supporting prior research findings and theoretical predictions, all resources included in the analysis correlated non-negatively with the proof validation performance (see Table 6, upper line). Except for basic procedural mathematical knowledge and metacognitive awareness, all resources showed a significant weak to moderate correlation.

Table 6. Correlations of cognitive resources and proof validation / proof construction performance.

	Cognitive Resources							
	BCMK	BPMK	ACMK	APMK	MSK	PSS	CRS	MA
Proof Validation	.503***	.167	NA	NA	.322**	.318**	.288*	.083
Proof Construction	NA	NA	.319**	.380***	.218**	-.002	.263*	.071

$^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$; NA: Not analyzed

The averaged Generalized Linear Mixed Model (GLMM) for proof validation, calculated to analyze the impact of the six resources on students' performance on proof validation (RQ1), can be found in Table 7 (upper part). It reveals that out of the six individual resources, only basic conceptual knowledge (β = .67; p = .016) and mathematical strategic knowledge (β = .54; p = .024) showed a significant influence on students' performance in proof validation. Furthermore, problem-solving skills (β = .47; p = .053) and metacognitive awareness (β = .37; p = .084) showed a relatively strong impact, although not reaching significance. Conditional reasoning skills were connected slightly negatively with proof validation performance when controlling for the other resources, and basic procedural mathematical knowledge showed only a weak connection.

Table 7. Regression coefficients of the individual cognitive resources for the optimal, averaged Generalized Linear Mixed Models for proof validation / construction.

	Cognitive Resources							
	BCMK	BPMK	ACMK	APMK	MSK	PSS	CRS	MA
Proof Validation	.67 .016	.07 .402	NA	NA	.54 .024	.47 .053	-.23 .197	.37 .084
Proof Construction	NA	NA	.30 .020	.39 .005	.23 .038	-.03 .427	.20 .068	.01 .454

5.2.7.3 The Influence of the Resources on Proof Construction Performance

Again, as expected all resources correlated non-negatively with the proof construction performance (see Table 6, lower line). Interestingly though, problem-solving skills did not show a significantly positive correlation with proof construction.

The calculated GLMM to analyze the impact of students' resources on their performance in proof construction (see Table 7, lower part) shows significant influences only by domain-specific resources: advanced procedural mathematical knowledge (β = .39; p = .005) showed the strongest relation to proof construction performance, followed by advanced conceptual mathematical knowledge (β = .30; p = .020) and mathematical strategic knowledge (β = .23; p = .038). Furthermore, students' conditional reasoning skills show a medium impact (β = .20; p = .068) not reaching significance.

5.2.7.4 Proof Validation as a Predictor of Proof Construction

To analyze the relation of proof validation and proof construction skills, we calculated Spearman's rank correlation coefficient between students' mean scores on proof validation and proof construction tasks. As expected, a significant weak positive correlation of $r_s(62) = .22$, p = .044 could be observed.

To analyze the predictive value of students' proof validation performance on their proof construction performance when controlling for the six cognitive resources, we calculated another GLMM including students' mean scores on proof validation tasks as an additional independent variable besides the six resources. The resulting model shows no significant influence of students' proof validation performance on the proof construction performance (β = .06; p = .389) beyond that of the resources. Including proof validation skills left the regression coefficients of the six resources largely unchanged (APMK: β = .39; p = .005, ACMK: β = .30; p = .021, MSK: β = .23; p = .039).

5.2.8 Discussion

Proof construction and proof validation are regarded as central learning goals and activities for mathematics students, especially at the university level. Yet, students of all ages have severe difficulties with both activities (e.g., Healy & Hoyles, 2000; A. Selden & Selden, 2013; Weber, 2003) and differ greatly in their performance. The current study examined to which extent these differences can be attributed to the availability of various individual cognitive resources underlying students' proof construction and proof validation performance (Blömeke et al., 2015). Several of such cognitive resources have been suggested to underlie students' mathematical argumentation and proof skills, and therefore also their proof validation and proof construction performance. Yet, their relative influence was rarely studied in the past. The reported study used multiple GLMMs to estimate the relative influence of each resource on beginning university students' proof validation and construction performance. Furthermore, it analyzed the role of proof validation for solving proof construction tasks.

In the following, we will first discuss the influence of the three domain- respectively content-specific resources and then focus on the domain-general resources. Subsequently their relation will be discussed and afterwards we will close with the general results of the study, its implications, and a short outlook.

Content- and domain-specific knowledge showed a strong impact on students' performance in proof validation and construction. Both analyses highlight the impact of these resources, which proved to be the only significant predictors of students' performance in our data. Their high influence is in line with prior research (e.g., Chinnappan et al., 2012; Ufer et al., 2008) and shows that handling mathematical proofs is a knowledge intensive activity. The results therefore also match current research on scientific reasoning in general, which underlines its domain-specificity (e.g., Kuhn, 2002; Schunn & Anderson, 1999; Zimmerman, 2000).

An interesting result from the employed GLMMs is the significant predictivity of *mathematical strategic knowledge* on the performance in proof validation and construction (β = .54 for proof validation; β = .23 for proof construction). This resource was introduced by Weber (2001), yet especially quantitative research related to this resource is scarce. Our study extends Weber's original work by proposing an instrument to survey students' mathematical strategic knowledge systematically. However, this is only a starting point for future assessment and research on this construct. Further research may comprise an improved theoretical analysis of the construct as well as its relations to existing knowledge concepts (de Jong & Ferguson-Hessler, 1996). Especially a clear separation from problem-solving skills and heuristics on the one side, mathematical content knowledge on the other side, and more general constructs like expertise on the third side are needed to purposefully include it in future studies. Theoretical insights could then be quantitatively underlined to offer the basis for a systematic validation of our instrument. However, based on the results of our study we propose to give mathematical strategic knowledge a more prominent place in research on mathematical argumentation and proof skills.

Domain-general resources showed no significant impact on both students' performance in proof validation and proof construction and results regarding their impact on either one differ between both models. This contrasts prior research: Two empirical studies (Chinnappan et al., 2012; Ufer et al., 2008) in the context of school geometry proofs have shown a substantial impact of *problem-solving skills* on proof construction tasks, while in our data the impact of problem-solving skills is low and insignificant. One possible explanation may be that the inclusion of mathematical strategic knowledge in the GLMM reduces the impact of problem-solving skills, as mathematical strategic knowledge can partially be seen as a domain-specific analogue of problem-solving heuristics. However, since mathematical strategic knowledge and problem-solving skills did not correlate significantly (see Table 5) and the correlational analysis (see Table 6) does not show a significant correlation between problem-solving skills and proof construction, this explanation has to be dismissed.

Alternatively, university students may not have required knowledge about general problem-solving heuristics for the proof construction tasks as they had domain-specific strategies from their mathematical strategic knowledge at hand to solve these. Thus, students may not have applied their general problem-solving skills in these tasks as they had proof-specific strategies at hand. In contrast, the 9th grade students in the studies by Chinnappan et al. (2012) and Ufer et al. (2008) as novices had to rely on their *weak problem-solving heuristics* (Chinnappan & Lawson, 1996) leading to the high impact of problem-solving skills in both studies.

A third explanation may be that the proof construction tasks centered on a different part of the problem-solving process than the employed problem-solving tasks: The proof construction tasks were chosen similar to common tasks from proof-based real analysis lectures so that participating students likely already knew how to handle tasks like these, could quickly structure their problem space, and had a set of reasonable operations at hand to perform on the task. The focus of the tasks, therefore, was on the purposeful combination of the operators to solve the problem and subsequent formal-deductive aspects of proof construction. In contrast, the problem-solving tasks were mostly unfamiliar to the students and required students to create and structure the problem space and identify reasonable operations. Once these were identified, the actual solution of the task was rather straightforward. This difference in both kinds of tasks may have led to the fact that the measured problem-solving skills had no significant impact on the proof construction performance.

Analogous to proof construction, problem-solving skills also had a non-significant impact on students' performance in proof validation, yet there is a significant correlation between problem-solving skills and students' performance in proof validation. As most students likely had never encountered explicit proof validation tasks in their lectures before, they may have had to invest time in structuring these validation problems and finding reasonable operations for refuting or accepting the given proofs, much like they had to for the problem-solving tasks.

Overall, further insights on the role of problem-solving strategies in proof construction and validation (focusing on various levels of expertise and different aspects of the problem-solving process) would be valuable. We advocate that it is important to consider which aspect of a proof construction or validation task leads to students' difficulties: Is it the creation and structuring of the problem space (which relies largely on content knowledge), the availability of permitted and effective operations within the problem space (which relies on mathematical strategic knowledge and/or problem-solving heuristics), the correct sequencing of operations within the problem space (which relies on self-regulation skills and metacognition), or is it the formal-deductive aspect of proof construction. Regarding the data from this study, especially the low impact of metacognitive awareness, the first two or last seem to have been the major aspects in the employed proof construction tasks.

In both models, *conditional reasoning skills* play a minor role compared to the domain-specific resources included. For proof validation, this may imply that students rarely check the deductive structure of a proof on a formal level. Rather they seem to concentrate on the evaluation of the task statements (conceptual knowledge) and on the general proof approach (mathematical strategic knowledge). For proof construction, this may imply that students do not create their proofs inference by inference, but rather focus on the semantics of an inference, for example by using mental models (Johnson-Laird, 1980; Ufer, Heinze, & Reiss, 2009a), or work with larger chunks like mimicking whole proof steps from already known proofs. Both strategies would then rely heavily on their mathematical knowledge, which corresponds to the results from the GLMMs.

The results regarding the relation of proof validation and proof construction are striking. First, a significant, positive correlation between students' proof validation and proof construction performance was found in our data. Although in this study the tasks for proof validation and proof construction were related to different content areas, the correlation between both equals other studies (Ufer, Heinze, Kuntze, et al., 2009) using only one content area, suggesting that proof validation may be a content-overarching skill. Yet, under control of the six resources, no impact of proof validation on proof construction could be shown. Thus, it appears that the correlation between both is not a sole effect of methodological knowledge (Heinze & Reiss, 2003), which is assumed to underlie both proof construction and proof validation, but rather an effect of the other included resources. Further, the assumption that proof validation as a sub-process of proof construction should play an important role for proof construction therefore is not supported by our data and it seems that the cognitive resources underlying both performances lead to the observed correlation. This is in line with our theoretical model (see Figure 3), suggesting that both proof construction and proof validation performance have the same underlying skill.

5.2.9 Limitations and Conclusions

The analyses employed in this study provide first insights regarding the impact of several cognitive resources on proof validation and proof construction performance and therefore on students' mathematical argumentation and proof skills. Still, the results need to be handled with care: First, the number of participants for the study is low. Although this limitation could be partially overcome by using GLMMs for the analysis, a replication of the study with more participants would be desirable. Second, there is not *one* way of model selection in GLMMs. Accordingly, different methods such as forward inclusion, backward elimination, AICc based selection, or averaging may lead to different models, each being *optimal* in a certain way. Still, we are confident that the approach taken here is reasonable as it is recommended by GLMM literature (Bolker et al., 2009; Zuur et al., 2009) and reflects the data well.
Third, the obtained results could naturally be dependent on the conceptualization and operationalization of the resources and on the content areas of the tasks used for proof validation and proof construction. Further research would be beneficial to ensure the generalizability of the results. Here foremost a thorough theoretical analysis of mathematical strategic knowledge, as well as replication studies, would be desirable.

Concluding, the study gives explicit results on the impact of the different individual cognitive resources on students' proof validation and construction performance. Domain- and content-specific resources such as conceptual mathematical knowledge, procedural mathematical knowledge, and mathematical strategic knowledge play a superior role in explaining students' performance, whereas domain-general resources seem to have a weaker influence. A straightforward implication for university teaching would therefore be to help students improve their domain-specific resources and not focus on domain-general or knowledge lean interventions, which show mixed results and are a matter of broad discussion (D. W. Eccles & Feltovich, 2008; Greiff et al., 2014; Sweller, 1990; Tricot & Sweller, 2014). Still, if this is possible, to which extent this is possible, and how this is possible, especially for the yet not overly explored mathematical strategic knowledge, cannot be answered by this study but has to be investigated in future intervention studies.
Overall, the results give a first insight on what counts to be successful in proof construction and proof validation at a university entry level and therefore promotes new ideas regarding the teaching of argumentation and proof at university level.

5.3 Instructional Approaches to Support Complex Cognitive Skills and Their Resources: Comparing the Effects of Supporting the Resources One-by-one or Simultaneously

5.3.1 Introduction

Today[23], educators in formal as well as informal learning settings are concerned with increasingly complex learning goals such as argumentation, information or complex problem solving, as well as other 21st century skills (e.g., Greiff et al., 2014; National Research Council, 2012; OECD, 2013; Osborne, 2013; Scherer, 2015; Trilling & Fadel, 2009). Accordingly, one major aim of interdisciplinary as well as disciplinary educational research is to provide practitioners, teachers, and lecturers with research results on how these learning goals can best be achieved and how students can be effectively supported in their learning attempts. In case of mathematical argumentation and proof skills, this aim is tackled internationally by multiple research groups (e.g., Andriessen, 2009; Conner, Singletary, Smith, Wagner, & Francisco, 2014; Fukawa-Connelly, 2014; Heinze & Reiss, 2009; Kollar et al., 2014).

Still, research so far often failed to acknowledge mathematical argumentation and proof skills, as well as other skills, as *complex cognitive skills*, that is having several underlying resources that need to be coordinated and integrated to solve given problems or meet certain situations requiring the complex cognitive skill. In the case of mathematical argumentation and proof skills, several underlying resources such as *mathematical content knowledge*, *methodological knowledge,* or *problem-solving skills* have been proposed by prior research (Heinze & Reiss, 2003; Schoenfeld, 1985) and their influence was pronounced on a theoretical level by process models and described by multiple studies (e.g., Boero, 1999; A. Selden & Selden, 2013): For example, students faced with a mathematical proof task require mathematical content knowledge to identify the objects within the task and unpack their definitions and meaning. As proof tasks represent problems for students, problem-solving skills are needed to guide students' search for a solution and to purposefully apply heuristics to construct a proof, which meets the acceptance criteria of the local mathematical community.

The resources of mathematical argumentation and proof skills further received at least partial empirical support (Chinnappan et al., 2012; Ufer et al., 2008), yet they are rarely considered comprehensively in current research (see section 5.1). That is, although researchers acknowledge that complex cognitive skills depend on several underlying resources such as knowledge facets (e.g., Blömeke et al., 2015; Schoenfeld, 2010; Shulman, 1987), these underlying resources are rarely considered in the design of learning environments.

Yet, acknowledging the resources underlying a complex cognitive skill leads to an instructional dilemma: Is it favorable to focus on the individual resources and support their acquisition and thereby indirectly the overarching skill? Or should the focus rather be on the complex cognitive skill, explicitly including all resources at once? Both approaches appear to have advantages: The first approach benefits from a higher instructional clarity as all resources are addressed individually, yet also requires the later transfer from the individual resources to the overall skill. In contrast, the second approach may be partially overwhelming students with the complex

[23] Parts of this study and preliminary analyses have been presented at the conference ICLS 2016 and published in the proceedings (Sommerhoff, Ufer, & Kollar, 2016a).

cognitive skill and its underlying resources all at once, yet allows an integrated learning of the resources in an authentic setting that automatically initiates the integration of the resources.
Accordingly, instructional approaches to support students in learning the resources underlying a complex cognitive skill alongside the complex cognitive skill itself need to be explored and contrasted regarding their effectiveness. The present study addresses this research gap and gives first insights on how this twofold aim of supporting the learning of the resources underlying a complex cognitive skill alongside the complex cognitive skill itself can be achieved.
Related research from instructional design has long been focusing on the support of complex cognitive skills, in particular part-task and whole-task strategies for learning (e.g., R. C. Anderson, 1968; Lim et al., 2009; Naylor & Briggs, 1963). Here, research contrasted two approaches: the *part-task approach*, which focused on the acquisition of individual sub-tasks or steps within a larger task to later integrate these into the whole task, and the *whole-task approach*, which focused on the immediate acquisition of the larger, whole task. One central tenet of that research was that complex cognitive skills are acquired more efficiently using a whole-task approach. This raises the question whether it is also more effective to support resources necessary to perform a complex cognitive skill simultaneously (analogue to the whole-task approach) or whether each should be developed separately one-by-one (analogue to the part-task approach). As the resources underlying a complex cognitive skill exceed individual steps or sub-tasks, may have to be purposefully applied within multiple steps, and require more than a sequential enchainment as compared to the individual part-tasks, the transfer of the central tenet from the part-task / whole-task debate is questionable.
In our study, we therefore contrast two instructional approaches to support the development of mathematical proof skills: A *one-by-one approach*, which focuses and supports each resource individually, and a *simultaneous approach*, which focuses and supports multiple resources in parallel, both using authentic tasks and settings for mathematical argumentation and proof activities. We compare students' learning outcomes resulting from the approaches, both on the individual resources as well as on their overall argumentation and proof skills to give first insights into the effects of both approaches and their feasibility in the context of mathematical argumentation and proof skills as well as more generally.

5.3.2 Background

5.3.2.1 Instructional Approaches for Complex Skills.

The idea that instructional strategies to support the learning of less complex skills may differ from those to support more complex skills has been raised repeatedly by educators and prior research (e.g., Branch & Merrill, 2011). Yet, the idea entails serious intricacies, starting with the notion of *skill complexity,* which is ill-defined. Furthermore, it is per se unclear which instructional approach would be suitable for which level or kind of skill complexity.
Naylor and Briggs (1963) gave a first account of *task difficulty*, differentiating two independent dimensions: *Task complexity*, accounting for the individual complexity of the sub-tasks, as well as *task organization*, describing the demands posed by the interrelationship between the various sub-tasks and their integration into the whole task. Their experimental study (Naylor & Briggs, 1963) suggests that tasks with a high sub-task complexity but low task organization benefit from part-task training. That is, skills for individual sub-tasks are trained, and afterwards connected using different sequencing strategies (e.g., forward chaining and snowballing). Contrary, tasks with low sub-task complexity but high task organization benefit from whole-task training. Further, tasks that require not only the execution and enchaining but also the integration of several sub-tasks can be more effectively taught using whole-task approaches.

Subsequent research contrived plausible theoretical arguments as well as empirical evidence for both approaches: Arguments for the part-task approach are mostly based on classical learning theories like ACT (J. R. Anderson, 1996) that assume the decomposability of complex skills into less complex part-skills (J. R. Anderson, 2002). This atomistic approach has been challenged by sociocultural and situated conceptions of learning that highlight the situatedness of learning (e.g., Brown et al., 1989; Greeno, 1998; Lave & Wenger, 1991; The Cognition and Technology Group At Vanderbilt, 1990). The whole-task approach also gained empirical support by evidence pointing to difficulties associated with attempts to transfer from part-tasks to the whole task (see J. R. Anderson et al., 1996 for a critical discussion). Fragmentation, compartmentalization, and the lack of transfer of learning (see further R. E. Clark & Estes, 1999; Perkins & Grotzer, 1997; van Merriënboer et al., 1997) are often mentioned as the pitfalls of part-task learning.

Today, several studies document the advantages of whole-task learning for a broad range of learning goals and many educational theories assume that learning is evoked and supported best by rich, meaningful tasks (van Merriënboer, 2002), which are hard to achieve by focusing solely on an atomistic approach dissecting whole tasks into its fragments.

However, recent empirical studies highlight that the benefit of focusing one’s attention may overweight the cost of later transferring and integrating the parts (So, Proctor, Dunston, & Wang, 2013) and that additional research is needed to identify which components and features of a whole skill influence how effective different learning strategies are (Lim et al., 2009; Wickens, Hutchins, Carolan, & Cumming, 2013). One such aspect determining the effectivity of instructional approaches appears to be prior knowledge or attainment (Salden, Paas, & van Merriënboer, 2006), as with low prior attainment both the part-tasks as well as their integration have to be learned.

Furthermore, compared to mathematical argumentation and proof skills, what has been described as „complex“ skills in earlier research (c.f., Gagné & Merrill, 1990; van Merriënboer, 1997; van Merriënboer, Kirschner, & Kester, 2003) seems to exhibit a lower complexity level. For example, creating spreadsheets for monthly sales figures (Merrill, 2002) or handling a mechanic excavator (So et al., 2013) cannot be seen as on the same level as argumentation skills, as here not only the integration of several sub-tasks or sub-skills in the sense of manual skills, operations, or activities is required, but rather the integration of various resources underlying the skill, which have to be monitored, coordinated, and regulated. Further, the resources have to be utilized in different ways, cannot be sequentially enchained, and have to be used concurrently, interacting with each other.

5.3.2.2 Complex Cognitive Skills and Underlying Resources

Complex cognitive skills are often conceptualized in the sense of Koeppen et al. (2008) as latent cognitive dispositions underlying a person’s performance in certain situations. For example, problem-solving skills refer to the cognitive disposition to succeed in various problem situations, that is situations in which an undesired initial state has to be transformed into a goal state, yet the needed operation to achieve this is not at hand (e.g., Dörner, 1979; Mayer & Wittrock, 2006, p. 287). Examples for different situations requiring problem-solving skills are analytical (assessed in PISA 2003), interactive (assessed in PISA 2012), or collaborative problem solving (assessed in PISA 2015) (Greiff, Holt, & Funke, 2013).

Similarly, mathematical argumentation and proof skills refer to the cognitive disposition to succeed in a set of proof-related situations and activities (Giaquinto, 2005; Mejía-Ramos & Inglis, 2009a; A. Selden & Selden, 2015a). Such situations may ask an individual to construct, that is to create a valid mathematical proof for a claim (e.g., A. Selden & Selden, 2015a), or validate mathematical proofs, that is to read a purported proof and judge its correctness. A person’s success in these situations will be judged relative to certain norms, which are shared within the local community and represent an important resource for mathematical argumentation and proof skills (e.g., Heinze & Reiss, 2003; A. Selden & Selden, 2015a; Ufer, Heinze, Kuntze, et al., 2009; Yackel & Cobb, 1996). This said, students' complex cognitive skills as a latent construct are not independent on their own. Several theoretical, as well as empirical accounts, underline that complex cognitive skills may require multiple underlying resources. For example, Shulman (1987) discusses several different knowledge facets (e.g., pedagogical knowledge, content knowledge) as underlying teaching skills and also problem-solving skills are assumed to have underlying resources, such as problem-solving heuristics (e.g., Abel, 2003; Schoenfeld, 1985). This kind of conception can also be found in vocational education, where Mulder et al. (2009, p. 757) speak of an "integrated set of capabilities consisting of clusters of knowledge, skills, and attitudes". Further, also the theoretical discussion and framework by Blömeke et al. (2015) integrate these ideas and conceptions, emphasizing the relations between underlying resources, complex cognitive skill, and task performance.

This conception of complex cognitive skills creates a situation, which is structurally similar to the part-task / whole-task debate. Here, students' complex cognitive skill (e.g., mathematical argumentation and proof skills) can be regarded as analogue to the whole-tasks, whereas the different resources underlying the complex cognitive skill (e.g., conceptual mathematical knowledge) are analogue to the part-tasks. This substantially broadens the part-task / whole-task debate, bringing up the question whether the results from the part-task / whole-task debate can be transferred. Here, the primary question will be, if the resources should be supported individually one-by-one or whether they benefit from a simultaneous approach. Further, the prior availability of the resources may affect the effectiveness of both approaches: Students with low prior attainment regarding the resources may benefit from the one-by-one approach, whereas those with a high prior attainment regarding the resources may benefit from the simultaneous approach. The answer to these question is highly relevant for the teaching and learning of any complex cognitive skill that is underpinned by several resources.

5.3.2.3 Mathematical Argumentation and Proof Skills and the Underlying Resources

Mathematics educators and educational psychologists widely agree that mathematical argumentation and proof skills can be seen as a complex cognitive skill, relying on several cognitive as well as non-cognitive resources. Over the last decades, several such resources have been proposed: They have been partly derived from models for more general skills like problem solving (Schoenfeld, 1985) or self-regulated learning (De Corte et al., 2000), been proposed by qualitative studies (Schoenfeld, 1987; Weber, 2001), and/or been partially empirically validated (Chinnappan et al., 2012; Ufer et al., 2008).

Based on these various frameworks and findings, the following four resources seem to comprise important cognitive resources for students' mathematical argumentation and proof skills (see section 5.2):

Mathematical content knowledge: One of the most fundamental and best-researched resources is mathematical content knowledge. Following widely accepted conceptions (e.g., J. R. Anderson, 1996; Hiebert, 1986; Star & Stylianides, 2013), it entails two facets, namely *conceptual mathematical knowledge (CMK)*, that is a network of knowledge about mathematical facts, theorems, objects, and their properties, as well as *procedural mathematical knowledge (PMK),* that is partly tacit knowledge, which is exercised in the accomplishment of a task (Hiebert & Lefevre, 1986). Both were shown to have a substantial impact on students' mathematical argumentation and proof skills (Chinnappan et al., 2012; Ufer et al., 2008). These results also match domain-general research on scientific reasoning underlining the importance of underlying, domain-specific knowledge (e.g., Kuhn, 2002; Schunn & Anderson, 1999; Zimmerman, 2000).

Mathematical strategic knowledge: In a qualitative study with mathematics students from different academic levels, Weber (2001) observed that mathematical content knowledge alone is not sufficient for being successful in constructing proofs. Despite knowing them in principle, students were often unable to identify concepts or methods necessary for a task or had problems applying them purposefully. As several studies have underlined this (e.g., Reiss & Heinze, 2004; A. Selden & Selden, 2013), it is assumed that students require *mathematical strategic knowledge (MSK)*, that is domain-specific strategies linking cues and hints within tasks with the mathematical methods and concepts that would be useful in the context of the task (Weber, 2001). In the broader context of research, mathematical strategic knowledge can be seen as a domain-specific version of general problem-solving heuristics.

Similar knowledge facets have already been studied in other contexts under the term *strategic knowledge*, for example regarding problem solving (Gok, 2010) or reading comprehension (Paris et al., 1983), or as *conditional knowledge* in artificial intelligence research (Lehmann & Magidor, 1992). The term is further also introduced in general conceptualizations of knowledge (e.g., de Jong & Ferguson-Hessler, 1996). Still, as the construct has so far received little attention in mathematics education research, a clear classification and in particular separation from these constructs (especially regarding problem solving) and other constructs such as experience or expertise is up to now unfortunately still lacking.

Methodological knowledge: Beyond content and mathematical strategic knowledge, also *methodological knowledge (MK)* (Heinze & Reiss, 2003; Ufer, Heinze, Kuntze, et al., 2009) is assumed to be an important resource underlying mathematical argumentation and proof skills. It comprises knowledge about acceptance criteria for mathematical proofs (e.g., the rejection of circular reasoning or the need for an explicit reference to an underlying theoretical background). Methodological knowledge is not only important when validating given proofs, but also plays an important role while constructing proofs as it involves those criteria the constructed proof is later checked against by the mathematical community. As methodological knowledge comprises knowledge about different types of proofs and their feasibility for certain proof tasks, it is essential for constructing a valid proof.

General problem-solving skills and heuristics: Next to these three rather domain-specific resources, the influence of *general problem-solving skills (PSS)* and *heuristics (PSH)* on proof construction has been noted several times (e.g., Polya, 1945; Schoenfeld, 1985). Both, *general problem-solving skills (PSS)* and *heuristics (PSH)*, are closely related and mostly conceptualized in a way that heuristics, that is rules-of-thumb for problem-solving processes, are employed when solving a problem and accordingly represent an important resource for problem-solving skills (e.g., Abel, 2003; Schoenfeld, 1985). Their importance is pronounced repeatedly (e.g., Reiss & Renkl, 2002; Schoenfeld, 1985) and also partially underpinned in the context of geometry proofs in secondary school contexts (Chinnappan et al., 2012; Ufer et al., 2008).

Proofs can be easily conceptualized as problems in the sense of Dörner (1979) or Schoenfeld (1985), as these are mostly non-routine tasks for students that they have no immediate solution strategy at hand for. The process of solving these problems can be regarded from an information processing perspective (Newell & Simon, 1972; Simon, 1978) and requires the creation of a problem space, entailing the initial state and the goal state, which have to be connected by using several operators. This is a multi-step search process that, if successful, generates a deductive chain of arguments as a solution for the problem (e.g., Heinze et al., 2008; Koichu & Leron, 2015; Weber, 2005).

Prior research has mostly underlined the importance of these four cognitive resources (Figure 42) for students' mathematical argumentation and proof skills. Mathematical content knowledge as well as problem-solving skills were shown to significantly explain the variance in students' mathematical proof construction (Chinnappan et al., 2012; Ufer et al., 2008) and also methodological knowledge shows a connection to their performance in proof construction (Ufer, Heinze, Kuntze, et al., 2009).
Still, the resources also represent important learning goals themselves, thereby creating a situation similar to the part-task / whole-task debate from instructional design. In this study, we aim to support the identified resources individually and by that indirectly also students' overall argumentation and proof skills leading to an improved proof construction performance.

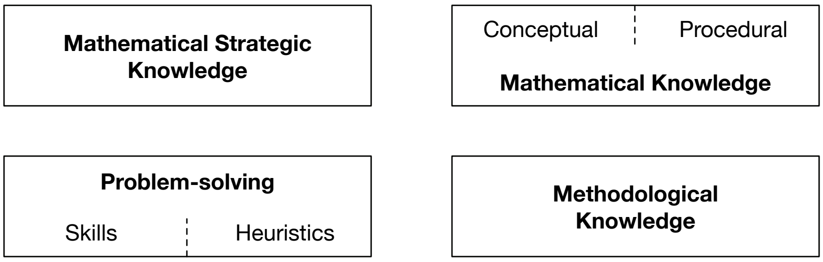


Figure 42. Four cognitive resources underlying mathematical argumentation and proof skills.

5.3.3 The Current Study

The present study is a first step to explore how acknowledging the resources underlying a complex cognitive skill can be functional in supporting the learning of the complex cognitive skill as well as its resources. The study therefore takes up the part-task / whole task debate from instructional design (J. R. Anderson et al., 1996; Branch & Merrill, 2011; Lim et al., 2009) in the pursuit of evidence for the feasibility and respective benefits of a *one-by-one* and *simultaneous* approach for supporting students' complex cognitive skill and its underlying resources.
This is done by examining students' mathematical argumentation and proof skills. These comprise a complex cognitive skill underpinned by several cognitive resources that have been identified in prior research. In a quasi-experimental study with university students, we investigated whether supporting each resource separately one-by-one or supporting all resources simultaneously yields higher learning gains on the resources as well as on overall mathematical argumentation and proof skills.
The research questions driving the study were:

RQ1 What are the effects of a one-by-one *vs.* a simultaneous approach for supporting overall mathematical argumentation and proof skills alongside its resources on the *resources*?

Our hypothesis was that the one-by-one approach would be superior in supporting the individual resources of a complex cognitive skill compared to the simultaneous approach. Each of the resources for mathematical argumentation and proof tasks as well as their utilization within argumentation and proof processes are already complex and shortcomings of students regarding prior knowledge, problem-solving skills, and other aspects have repeatedly been reported (e.g., Harel & Sowder, 1998; OECD, 2014; Reiss & Ufer, 2009; Schoenfeld, 1989; A. Selden, 2011). Thus, the results by Naylor and Briggs (1963) imply that a one-by-one approach should be better suited. Following the argumentation by Blömeke et al. (2015) we further assumed that instructional clarity would be higher in the one-by-one condition as each resource was covered individually. As multiple studies document that mathematical argumentation and proof tasks are complex for students (e.g., A. Selden & Selden, 1978; A. Selden & Selden, 2008, 2012; Weber, 2003), we envisioned that the higher instructional clarity would be helpful for the improvement of students' resources.

RQ2 What are the effects of a one-by-one *vs.* a simultaneous approach for supporting overall mathematical argumentation and proof skills alongside its resources on the *overall mathematical argumentation and proof skills*?

We assumed that the simultaneous approach would yield higher or at least comparable learning gains compared to the one-by-one approach. This is implied by the results from Naylor and Briggs (1963) as overall mathematical argumentation and proof skills require a high degree of "task organization", that is the underlying resources need to be purposefully combined and used. Accordingly, an approach integrating the resources and thereby allowing students to directly experience the simultaneous coordination of the resources could be favorable and lead to integrated learning. This is further supported by a prior review on part-task practice (Wickens et al., 2013) that revealed negative effects of part-task training when parts have to be used concurrently, which is the case with the resources underlying mathematical argumentation and proof skills.

Furthermore, the one-by-one approach requires students to later, that is after learning about each resource, integrate the various resources and their use when constructing mathematical proofs. As this does not arise as naturally as in the simultaneous approach, where the resources are already used in an integrated way, this should pose another obstacle for students from the one-by-one approach. In line with this argumentation, also situated learning theories (Brown et al., 1989; Lave & Wenger, 1991) suggest that students should benefit from the authentic, meaningful combination of resources as opposed to addressing them individually without the connections, for example between mathematical strategic knowledge and problem-solving skills.

5.3.4 Method

5.3.4.1 Design and Participants

To answer the research questions, we adopted a quasi-experimental research design with pre and post measurement, featuring two conditions corresponding to the one-by-one and simultaneous approach. The intervention was offered as a voluntary course for mathematics university students entitled "Mathematical proofs: That's how to do it!", which was aimed at undergraduate students after the first semester. A total of 45 students (18 male, 27 female, m_{age} = 20.82) participated in the study. Among them were 36 first year and 9 second year students who were either enrolled in a (financial) mathematics bachelor's program or a teaching degree for secondary education. Independent of the degree program, all participants can be assumed to have participated in proof-based real analysis lectures, giving the students the necessary foundation for the course. Of these 45 students, 21 students participated in the one-by-one condition whereas 24 students participated in the simultaneous condition of the course. Participants' mean university entrance grade (*M* = 1.92[24], *SD* = 0.52), as well as their final high-school grade in mathematics (*M* = 1.86, *SD* = 0.56), were in-between the best and second-best grade.

[24] Grades are scaled from 1 to 6, with 1 being the best grade.

5.3.4.2 Procedure

The course was scheduled across three consecutive days after the end of the regular lecture period and consisted of four two-hour intervention sessions plus two sessions for pretest and posttest. Without being aware of the difference, participants could choose to participate in one of both parallel groups, each representing one of the instructional conditions, that is the one-by-one approach or the simultaneous approach. The course was conducted by two experienced instructors with a mathematics as well as mathematics education background. Instructors swapped courses in the middle of the intervention to eliminate instructor effects.

The content of the course was based on the topics and proofs in *proof-based real analysis*, an introductory topic in university mathematics that is intensely covered during the first semester. Both courses covered the same teacher input, content, tasks, and time on task. The only difference was that tasks and content were arranged in a different order.

For the design of the intervention and the arrangement of the topics and tasks, we structured both conditions according to two levels within the course: the *macro-level* and the *micro-level*. The macro-level refers to the arrangement of the resources within the course, that is at which time each of the resources was addressed during the study and whether this was done in a one-by-one or a simultaneous manner. In contrast, the micro-level refers to instruction for each of the four resources themselves (Figure 43).

For the teaching of the individual resources, that is the micro-level, we adopted a 4C/ID inspired instructional design (van Merriënboer, 2013; van Merriënboer & Kirschner, 2007). This was done for two reasons: First, each of the resources is characterized by a lower *task organization*, that is the organization of aspects within the resources require less organization as compared to complete mathematical argumentation and proof tasks and therefore should benefit from a rather comprehensive instructional approach (Naylor & Briggs, 1963). Second, we wanted to parallelize the micro-level for both conditions as we are interested in the effects of the macro-level and differing instruction on the micro-level may have biased the results.

The employed micro-level design for teaching each of the resources consisted of an initial input phase with information on the resources, to support the resources directly. This was combined with a short list of elaboration and monitoring prompts (e.g., "Excerpt all important objects and properties from the task, explain these in your own words, and compare them to the formal definition.", "Search the task for keywords that you know from other tasks. What methods did you use there?"). The prompts were meant as procedural information while solving the proof construction tasks later. They were intended to scaffold the use and application of the individual resources during these tasks, to enhance students' analysis of the task according to each resource, and to induce students to elaborate and reflect on their knowledge regarding each resource. The prompts, therefore, ensured an explicit usage of the resources and hindered students to fall back into their old proof construction behavior. To show how these prompts could be purposefully applied, the instructor demonstrated the usage of the prompts.

5.3.4.2.1 The One-by-one Condition

The one-by-one condition was intended to support the different resources individually. Accordingly, the course was split into four individual blocks of two hours on the macro-level. During each of these four sessions, only one of the four resources was focused explicitly (Figure 43, upper part).

During each session, students worked on exactly four tasks and analyzed them focusing on the one resource that was covered during that session. Each task was then picked up in a second session and analyzed regarding the resource focused in that session. After that, the task was solved and discussed together with the instructor.

During the analysis process of the students and their work on the tasks, the instructors gave guidance, provided procedural information, and hinted students to use specific prompts from the provided list.

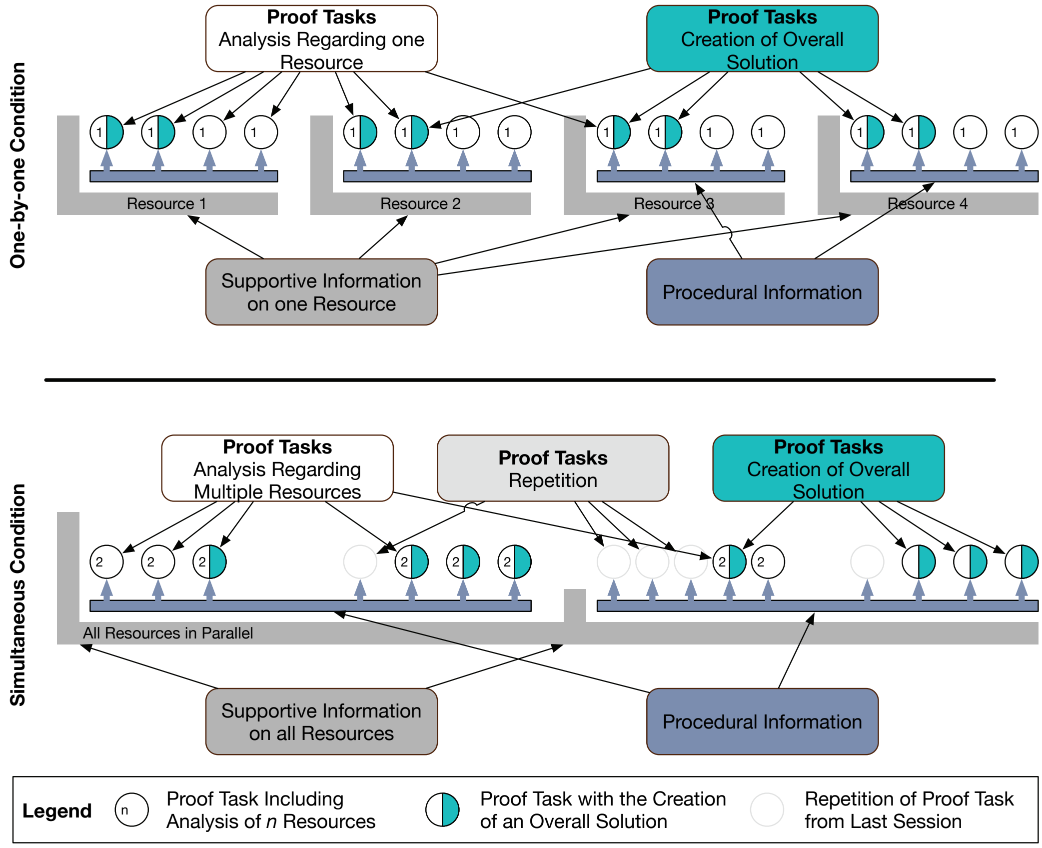


Figure 43. Instructional design used for both conditions within the intervention.

5.3.4.2.2 The Simultaneous Condition

Regarding the macro-level, the simultaneous condition included all four resources during each session, providing students with the opportunity to integrate the individual resources and see connections among them. That is, throughout each session, students had to focus on mathematical content knowledge, mathematical strategic knowledge, methodological knowledge, as well as problem-solving skills and heuristics while working on the tasks.

Here, input phases were given in the first and third session. As all resources were treated during each session, it was necessary to give a basic amount of supportive information on all resources in the first session, so that students would be able to work purposefully with all for resources. The remaining information was then introduced at the beginning of the third session.

Throughout the course, students from this condition worked on same eight proof tasks as those in the one-by-one condition, yet always analyzed them regarding two of the resources at once. The tasks were distributed over the sessions so that each resource would be covered in every session and each combination of two resources would occur equally often. The tasks that had already been analyzed and solved were reconsidered briefly in the next session so that each student worked on each task twice as in the one-by-one condition.

The students from the simultaneous condition received the same amount and kind of guidance as the students in the one-by-one condition.

5.3.4.3 Instruments

Pretest and posttest of the study included scales for each of the resources, one for students' argumentation and proof skills, as well as for covariates and demographic data. Most of the employed scales were adapted to the content, translated from English, or self-created if no suitable published scales were available in the literature. We mostly used parallelized tasks for the pre- and posttest to avoid repetition effects. We preferred this approach over using identical tasks, as it was especially important for the items within the *problem solving* and the *mathematical argumentation and proof* scale to be unknown and therefore retain a problem character (e.g., Dörner, 1979; Schoenfeld, 1985). Only for the problem-solving heuristics the same items were employed in pre- and posttest.

The employed scales had been piloted and evaluated prior to the reported study. Their reliability was $.58 < \alpha < .81$, with .58 corresponding to the only scale below .6 (mathematical strategic knowledge) that had been assessed using only four items. As a newly developed scale for a construct that has not been assessed quantitatively before, we decided to retain the scale despite of the low reliability.

The scales contained open as well as closed items. Closed items were evaluated using mark-recognition software with a subsequent manual control. The open items were coded by two raters following theory-based coding schemes. Double coding of over 15% of the data led to an interrater reliability of $\kappa > .78$ ($M = .93$; $SD = .10$). For each scale, sum-scores were calculated and scaled to values between 0 (worst) and 1 (best). Only the scale for problem-solving heuristics, which used Likert-type items, remained on a scale from 1 (worst) to 4 (best).

5.3.4.3.1 Dependent Variables

Mathematical content knowledge: The scales for conceptual (8 items) and procedural mathematical knowledge (5 items) were adapted from existing tests in the context of university mathematics (Rach & Heinze, 2016; Wagner, 2011) and slightly modified to fit the content area of the study. The conceptual items focused on assessing fundamental knowledge such as definitions, theorems, and properties of objects as well as their connections. The procedural items focused on routine procedures as solving equations or using the formula for the geometric sum, which were required in the employed proofs throughout the course and the according scales.

Methodological knowledge: The scale for students' methodological knowledge was based on existing scales from secondary school contexts (Heinze & Reiss, 2003; Ufer, Heinze, Kuntze, et al., 2009). Students' judgments regarding the validity of purported proofs were used as an indicator of their methodological knowledge.

Mathematical strategic knowledge: Mathematical strategic knowledge has to our knowledge not been quantitatively measured up to now. Building on the definition of the construct, we chose four typical tasks from the real analysis context as the foundation for four items. These were presented to students alongside four excerpts of the same task description. In a multiple-choice format students were then asked to select those task descriptions that hint towards a certain concept or method that would be helpful to solve the task. In a subsequent open question, they were asked to explain their choice and describe what the excerpts would imply. Closed and open item for each task description were combined using a dichotomous consistency rating, evaluating whether the selected excerpts combined with the given explanation matched the given task.

Problem-solving heuristics: To measure students' knowledge about and use of problem-solving heuristics, students were asked how often they made use of twelve different, prototypical problem-solving strategies (e.g., means-end analysis, or creating a sketch) taken from literature (Polya, 1945). Each of the strategies was reflected in four Likert-Scale items.

Problem-solving skills: Students' problem-solving skills were measured using four open items, asking students to solve problems that did not require domain-specific knowledge (neither mathematical nor from another domain), except for everyday knowledge and basic arithmetic skills. The items were then scored on a scale from 0 to 4, evaluating if the main steps for solving the problem were given and justified adequately.

Mathematical argumentation and proof skills: Besides the resources, a scale for assessing students' mathematical argumentation and proof skills consisting of four proof construction items was included. The tasks were chosen to be novel to the students, yet reflect prototypical tasks from real analysis lectures as well as those used within the intervention itself. The items were scored on a scale from 0 to 4, evaluating if the main ideas needed for a valid proof were given and adequately justified.

5.3.4.3.2 Further Variables

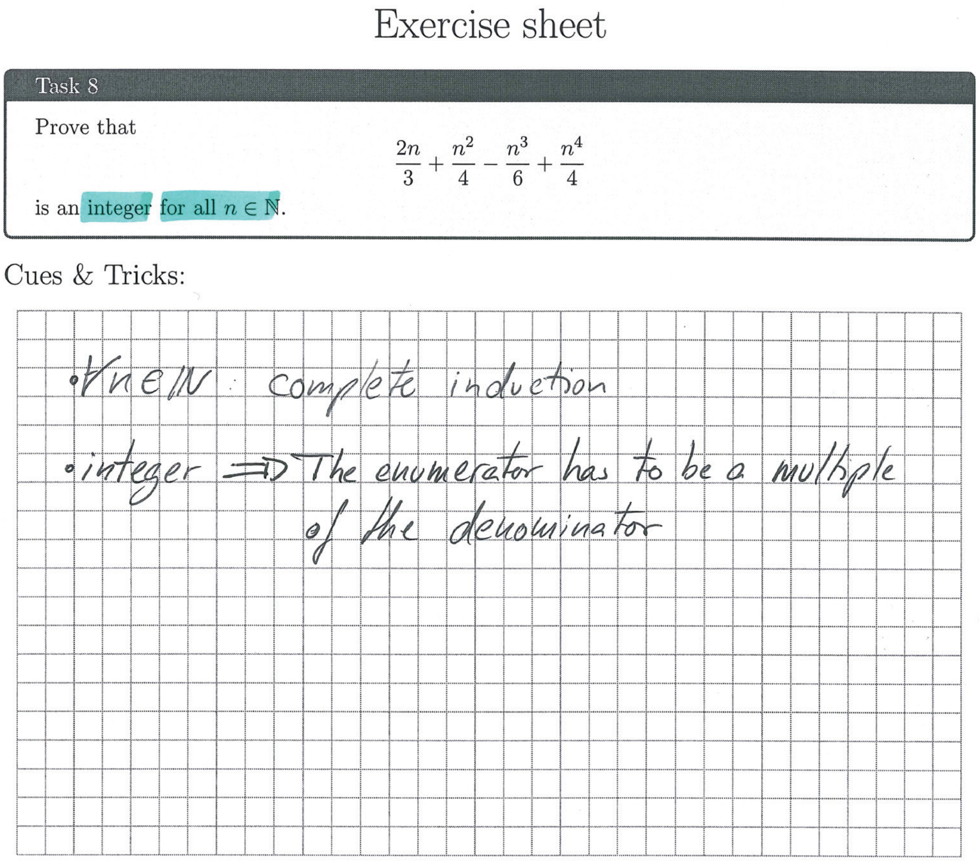
Besides the scales for the dependent variables, a scale for *conditional reasoning skills* from literature (Evans et al., 1995; Inglis & Simpson, 2008, 2009) with 16 items was included. As conditional reasoning skills are considered to be fundamental for any kind of reasoning activity, important for scholarly activities across disciplines, and were also shown to significantly predict certain aspects of mathematical argumentation and proof skills (Alcock, Bailey, et al., 2014; Kuhn, 2009; Leighton, 2006; Leighton & Sternberg, 2004), they were included in order to be used as a covariate in the later comparisons between conditions.

Furthermore, demographic data including gender, degree program, university-entry grade, and final high-school mathematics grade were gathered.

5.3.4.4 Implementation Check and Process Data

To check the implementation within both conditions and to create process data, students received prefabricated exercise sheets to work with for all tasks and analyses. The sheets were gathered and digitalized after every session throughout the intervention (see Figure 44 for an excerpt of an exercise sheet showing the analysis of a task regarding mathematical strategic knowledge). Subsequently, it was checked, whether students had explicitly analyzed the task regarding the resources and whether the analysis was done on a meaningful or a superficial level (dichotomous coding).

Additionally, a reflection scale on the content covered by the course was created for the posttest, probing students about several topics that may or may not have been covered by the course (e.g., "I think I learned a lot regarding problem solving"). To check that both conditions did indeed convey a one-by-one respectively simultaneous conception of the resources, students were also asked how separated they perceived the different resources during the intervention ("I think the course separated the individual prerequisites of proving well").


Exercise sheet

Task 8

Prove that

$$\frac{2n}{3} + \frac{n^2}{4} - \frac{n^3}{6} + \frac{n^4}{4}$$

is an integer for all $n \in \mathbb{N}$.

Cues & Tricks:

- $\forall n \in \mathbb{N}$: complete induction
- integer $\Rightarrow$ The enumerator has to be a multiple of the denominator

Figure 44. Excerpt of an exercise sheet of a student showing an analysis regarding mathematical strategic knowledge (translated).

5.3.4.5 Statistical Analysis

To answer the first research question, a MANCOVA using the resources as dependent variables and conditional reasoning skills as the covariate were employed to examine overall differences between both approaches. In line with this, effects of both approaches on individual resources were subsequently compared using an ANCOVA with conditional reasoning skills as the covariate. In the case of significant differences, the results were safeguarded by a further ANCOVA additionally including the pretest results of the resource.

The second research question was examined first by using an ANCOVA with students' mathematical argumentation and proof skills as dependent variable and conditional reasoning skills as covariate, then backing up the analysis with another ANCOVA including the pretest results as an additional covariate.

Throughout the analyses the significance level was set to 5%.

5.3.5 Results

5.3.5.1 Treatment Check

A qualitative analysis of the documents used throughout the intervention confirmed that students in both conditions actively analyzed the tasks regarding the respective resources before solving them and used the provided reflection prompts. Overall, 92.5% (92.6% in one-by-one; 92.4% in simultaneous) of the suggested analyses regarding the resources were done by the students, 1.9% were completely missing, and 5.6% of the analyses were on a superficial or non-purposeful level. This indication of a correct implementation of both conditions is further supported by the results of the posttest: A related samples Friedman two-way analysis of variance by ranks on the reflection scale regarding the covered topics used in the posttest

showed overall significant differences between students' answers on the covered topics ($\chi^2(6) = 89.048$, $p < .001$). Post-hoc Dunn-Bonferroni tests showed significantly lower values for both topics not covered during the course (beliefs, quantifier logic) in comparison to those covered by the course, indicating that students had paid attention throughout the intervention and were able to identify the resources they had worked on.
Furthermore, a Mann-Whitney U test on the question regarding the perceived separateness of the resources showed the expected significant difference ($U = 327.0$, $p = .029$; $M_{\text{one-by-one}} = 3.0$ & $M_{\text{simultaneous}} = 3.3$) between both experimental conditions, indicating that the participants of the one-by-one condition perceived the resources as more separated than the students from the simultaneous condition.

5.3.5.2 Preliminary Analyses

The descriptive results of the employed scales in the pre- and posttest (Table 8) showed acceptable values. No signs of floor or ceiling effects could be determined and the resulting variances were acceptable, too. The distributions of the resources and mathematical argumentation and proof skills within each condition were analyzed regarding normality and equality of variances, which they passed (Field, 2009; Gravetter & Wallnau, 2016).
To safeguard against possible problems regarding the comparability of the parallelized pre- and posttest scales, Pearson correlations for each pair of parallelized scales (CMK, PMK, MK, PSS, MA&P) were calculated, showing highly significant correlations ($r(43) = .447 - .669$, $p \leq .001$).

Table 8. Mean values for the scales obtained for both conditions in pre- and posttest.

	One-by-One				**Simultaneous**			
	Pretest		Posttest		Pretest		Posttest	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Conceptual Mathematical Knowledge	.29	.20	.38	.24	.36	.19	.44	.15
Procedural Mathematical Knowledge	.36	.18	.51	.24	.45	.19	.54	.20
Methodological Knowledge	.40	.16	.54	.14	.49	.17	.55	.16
Mathematical Strategic Knowledge	.35	.16	.57	.16	.39	.17	.69	.18
Problem-solving Heuristics	2.90	.21	3.13	.22	2.95	.23	3.20	.22
Problem-solving Skills	.43	.19	.35	.18	.44	.20	.40	.18
Mathematical Argumentation and Proof Skills	.34	.14	.29	.14	.36	.18	.32	.14

The results of the pretest regarding the dependent variables, that is the resources as well as students' mathematical argumentation and proof skills, suggest that both conditions were comparable prior to the intervention (Table 8). This was confirmed by calculating independent samples *t*-Tests comparing these variables between both conditions. None of the tests gained significance ($t(43) < 1.622$, $p > .112$), solely methodological knowledge slightly approached significance ($t(43) = 1.745$, $p = .088$) in favor of the participants in the simultaneous condition.

The same insignificant differences were found for the gathered demographic data and for students' conditional reasoning skills ($t(43) = .728$, $p = .361$).

5.3.5.3 Effects on the Resources

The descriptive results of the posttest (Table 8) show learning gains within both conditions for most resources, leading to pre-posttest effect sizes of $g = .35$ - 1.76 (Table 9). Solely students' problem-solving skills showed no gains, but rather decreased slightly ($g_{\text{one-by-one}} = -0.40$ & $g_{\text{simultaneous}} = -0.17$).

Table 9. Longitudinal effect sizes for both conditions.

	CMK	**PMK**	**MK**	**MSK**	**PSH**	**PSS**	**MA&P**
One-by-one	0.40	0.73	0.91	1.39	1.06	-0.40	-0.31
Simultaneous	0.47	0.50	0.35	1.76	1.07	-0.17	-0.28

Comparing the descriptive results of the posttest between both conditions (Table 8), slightly higher mean scores for all resources within the simultaneous condition as compared to the one-by-one condition can be observed. A one-way MANCOVA of students' posttest results on the resources (CMK, PMK, MK, MSK, PSH, PSS) using students' conditional reasoning skills as a covariate and both conditions as a factor showed no significant differences between both conditions ($V = .180$, $F(6,37) = 1.353$, $p = .259$, Pillai's trace). However, separate univariate ANCOVAs on the dependent variables revealed a significant difference between both conditions on mathematical strategic knowledge ($F(1,42) = 5.682$, $p = .022$, $\eta^2 = .119$) while controlling for conditional reasoning skills. All other ANCOVAs were insignificant ($F(1,42) < 1.149$, $p > .290$).
To validate this finding, we calculated another ANCOVA for students' post-test results on mathematical strategic knowledge, this time also including students' pretest results for mathematical strategic knowledge as a covariate. The according ANCOVA revealed significant differences in students' mathematical strategic knowledge between both conditions ($F(1,41) = 5.190$, $p = .028$), yielding a medium effect size ($\eta^2 = 0.112$) in favor for the simultaneous condition.
The results of the ANCOVA are also reflected when examining the learning gains of both conditions regarding mathematical strategic knowledge: They show significant longitudinal learning gains (paired samples t-Tests: one-by-one: $t(20) = -10.192$, $p < .001$; simultaneous: $t(23) = -7.478$, $p < .001$) from pretest to posttest, yet the effect in the simultaneous condition is larger ($g_{\text{one-by-one}} = 1.39$ & $g_{\text{simultaneous}} = 1.76$; Figure 45, left side).

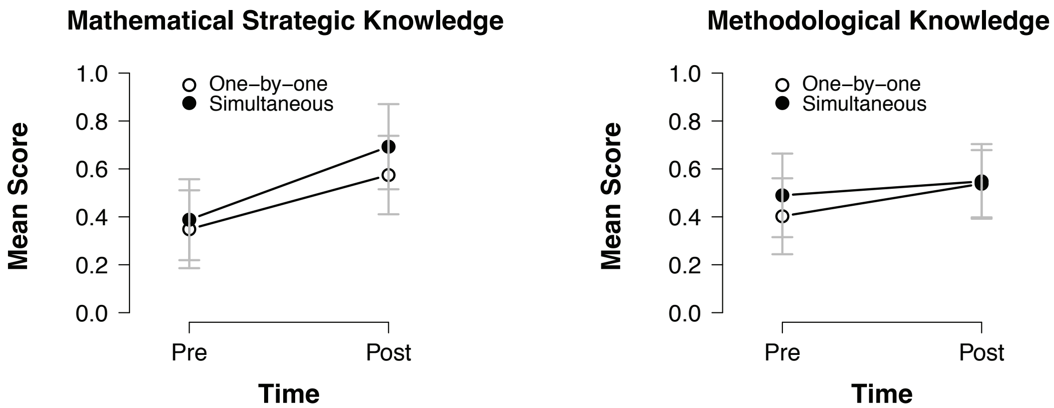


Figure 45. Effects of both approaches on mathematical strategic knowledge (left) and methodological knowledge (right).

Although not reaching statistical significance, the data gives a first indication for a between-conditions effect of methodological knowledge (Figure 45, right side; $F(1,41) = 0.428$, $p = .517$). The graph shows that the gains in the one-by-one condition ($g_{\text{one-by-one}} = 0.91$; $t(20) = -4.238$, $p < .001$) are larger than in the simultaneous condition ($g_{\text{simultaneous}} = 0.35$; $t(23) = -1.664$, $p = .110$), indicating that students in the one-by-one condition caught up with the students from the simultaneous condition.

5.3.5.4 Effect on Students' Argumentation and Proof Skills

The descriptive results of the pretest and posttest for students' mathematical argumentation and proof skills (see Table 8) and the according longitudinal effect sizes in both conditions ($g_{\text{one-by-one}} = -0.31$ & $g_{\text{simultaneous}} = -0.28$) show a slight decrease of students' performance, possibly due to after all more difficult posttest items. A one-way ANCOVA on students' mathematical argumentation and proof skills in the posttest, controlling for students' conditional reasoning skills, showed no significant difference ($F(1,42) = 0.337$, $p = .564$) between both conditions. This is underlined by the analogous ANCOVA, additionally controlling for students' results on mathematical argumentation and proof skills from the pretest ($F(1,41) = 0.144$, $p = .706$).
To further examine these effects, we performed an exploratory analysis to compare the effects of both conditions on students with different prior attainment, as prior research implied according differences. For this, median-splits according to students' pretest results on mathematical argumentation and proof skills were calculated, making it possible to examine the effects of both approaches on initially weaker respectively stronger students. The split resulted in four groups, a weaker and a stronger group for both instructional approaches. Calculating the longitudinal effects on the four groups showed mixed effects of the intervention (Table 10).

Table 10. Longitudinal effect sizes on students' mathematical argumentation and proof skills for the median-split groups.

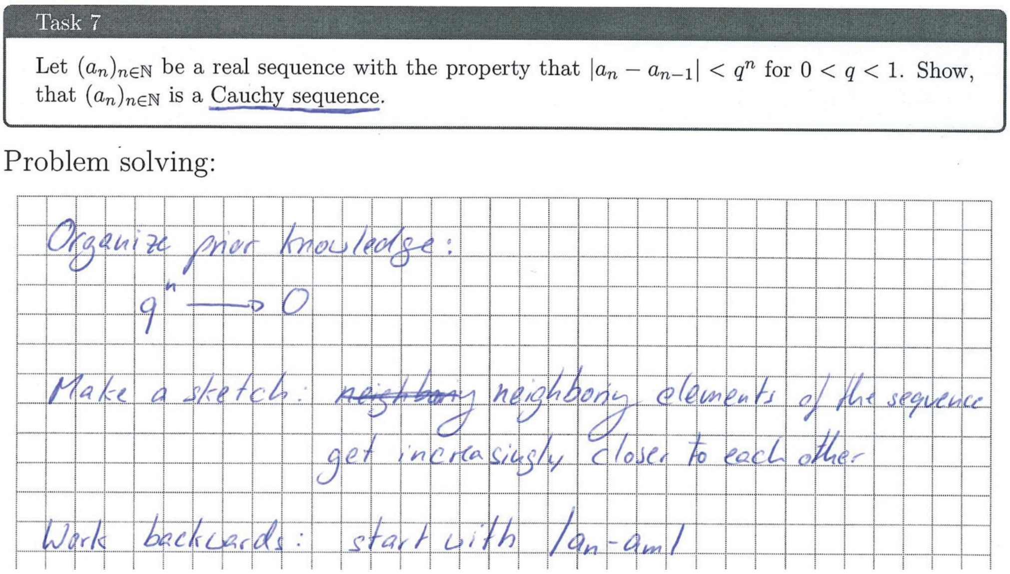
		Number of Students	**Pretest**		**Posttest**		**Effect Size *g***
			M	*SD*	*M*	*SD*	
Weaker	One-by-one	11	.22	.09	.22	.10	-0.06
	Simultaneous	8	.15	.11	.23	.10	0.74
Stronger	One-by-one	10	.46	.06	.38	.14	-0.81
	Simultaneous	16	.47	.08	.36	.14	-0.93

Apparently, students with higher initial mathematical argumentation and proof skills did not benefit from both approaches. On the other hand, results for the initially weaker students showed that the simultaneous conditions lead to substantial gains, whereas the one-by-one condition led to a negligible change. Although group sizes are small, the negative effect on the stronger, simultaneous group was significant ($t(15) = 3.137$, $p = .007$), yet the other effects did not reach significance ($p > .072$).

5.3.5.5 The Simultaneous Condition – an Illustration of the Effects

The exploratory analysis revealed first signs of an expertise reversal effect, that is initially strong students, which appear to have more prior knowledge on the resources and connections between those, benefit less from the intervention as compared to initially weaker students. Here, especially weaker students in the simultaneous condition seem to benefit from the intervention, as the concurrent focus on multiple resources seems to lead to a better integration and handling of argumentation and proof tasks. For the weaker students, the integration of the resources therefore seems to be particularly beneficial.

Even though these results cannot be further underpinned statistically, a qualitative examination may provide insights into the effects of the simultaneous approach for students with low initial argumentation and proof skills. For this purpose, we provide a deeper analysis of Leia, a student from the "weaker – simultaneous" group. Leia is 23 years old, in the first year of her bachelor mathematics studies. She failed both exams from the first semester, which draw heavily on proof construction.


Task 7

Let $(a_n)_{n\in\mathbb{N}}$ be a real sequence with the property that $|a_n - a_{n-1}| < q^n$ for $0 < q < 1$. Show, that $(a_n)_{n\in\mathbb{N}}$ is a Cauchy sequence.

Problem solving:

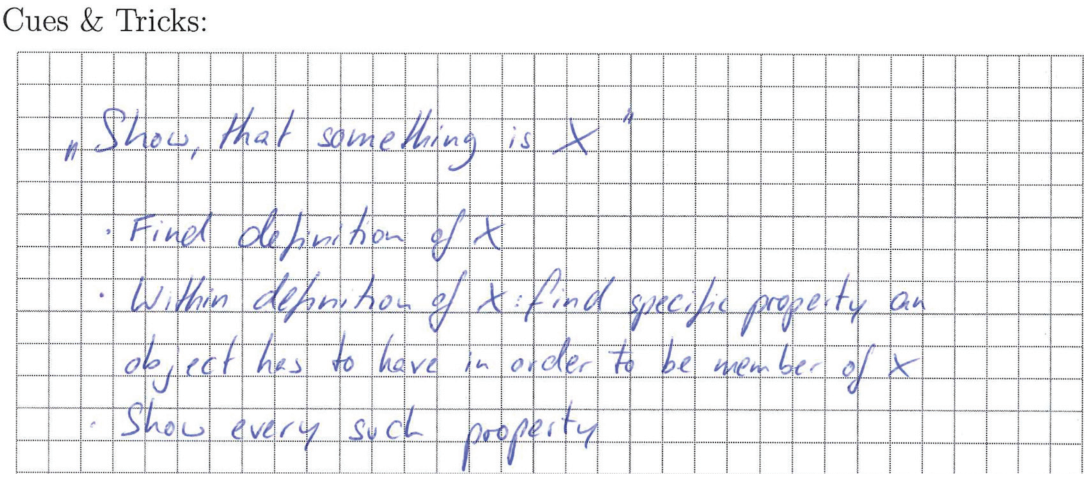
Organize prior knowledge:
$q^n \longrightarrow 0$

Make a sketch: ~~neighbory~~ neighboring elements of the sequence get increasingly closer to each other

Work backwards: start with $|a_n - a_m|$

Figure 46. Task description and Leia's bullet points regarding problem solving (translated).

Leia's work on the analysis of the given proof task regarding the resource *problem solving* (Figure 46) shows three main thoughts, each fitting to one of the elaboration and monitoring prompts given to the students. The first two mirror her attempts to recall the meaning of the property of the given sequence and to make sense of it, which seems to work out to a certain degree as the second point correctly reflects the given property. The third point then shows that she has created a plan for solving the task, even before actively trying to do so in her actual proof attempt. That is, she plans to use the general *problem-solving heuristic* of working backwards, here starting from the defining property of a Cauchy sequence (given in mathematical notation). This strategy matches her work regarding the *mathematical strategic aspects* (Figure 47). By concentrating on the structural parts of the given task, Leia unveils its type, referring to it as a "Show, that something is X"-task. She then lays out a broad idea on how to solve this type of task, by finding the properties that have to hold for an object to be a member of the class "X" and then showing that these properties hold. This mirrors the strategy of working backwards mentioned in the problem-solving analysis, which starts exactly with these properties.



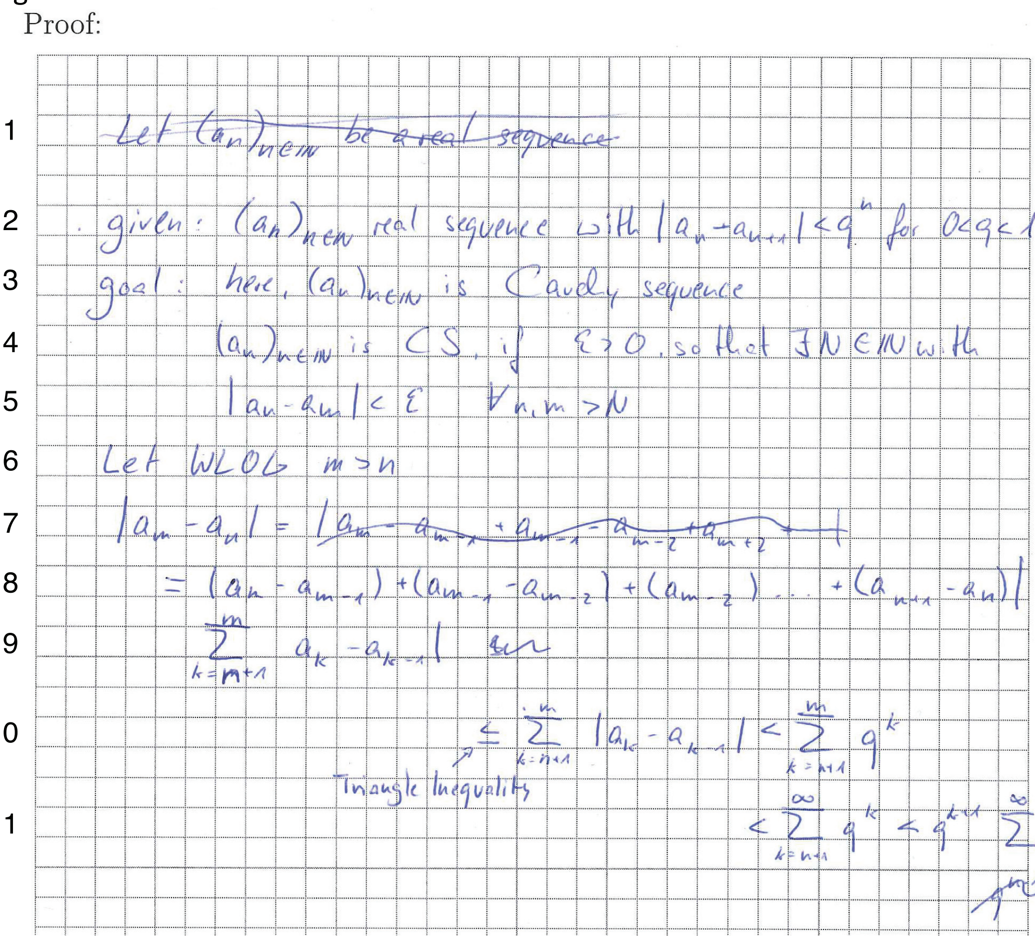
Cues & Tricks:

„Show, that something is X"

- Find definition of X
- Within definition of X: find specific property an object has to have in order to be member of X
- Show every such property

Figure 47. Leia's bullet points regarding mathematical strategic knowledge (translated).

After carrying out both analyses, Leia starts her proof attempt (Figure 48). Apparently, she jumps quickly into the proof, but is unhappy with her first approach and crosses it out (line 1). As the crossed-out line is correct and resembles a reasonable approach for the task, it can be assumed that Leia hesitates because she wants to stick to the information and procedures given to her in the intervention, asking her to clearly clarify what is given and, in particular, her goal. Starting in her third line, she then lays out the definition of a Cauchy sequence (with one minor error in line 4), which she then uses in her actual proof attempt, starting from line 6. Here, she can successfully reduce the property of a Cauchy sequence to the property of the given sequence (line 8-10) but then fails to explicate the last proof step and conclude that the resulting term converges to zero as n increases.



Proof:

1 ~~Let $(a_n)_{n\in\mathbb{N}}$ be a real sequence~~

2 given: $(a_n)_{n\in\mathbb{N}}$ real sequence with $|a_n - a_{n+1}| < q^n$ for $0<q<1$

3 goal: here, $(a_n)_{n\in\mathbb{N}}$ is Cauchy sequence

4 $(a_n)_{n\in\mathbb{N}}$ is CS, if $\varepsilon > 0$, so that $\exists N \in \mathbb{N}$ with

5 $|a_n - a_m| < \varepsilon \quad \forall n,m > N$

6 Let WLOG $m > n$

7 $|a_m - a_n| = |\cancel{a_m - a_{m-1} + a_{m-1} - a_{m-2} + a_{m+2}}|$

8 $= |(a_m - a_{m-1}) + (a_{m-1} - a_{m-2}) + (a_{m-2}) \ldots + (a_{n+1} - a_n)|$

9 $\sum_{k=m+1}^{m} a_k - a_{k-1}|$

10 $\leq \sum_{k=n+1}^{m} |a_k - a_{k-1}| < \sum_{k=n+1}^{m} q^k$

Triangle Inequality

11 $< \sum_{k=n+1}^{\infty} q^k < q^{k+1} \sum^{\infty} q^k$

Figure 48. Leia's proof attempt (translated; line numbers added).

Leia's work indicates that especially low attaining students may benefit from a structured approach to mathematical argumentation and proof tasks. In her case, the explicit discussion of aspects of the task related to the resources needed for the task, helped her to plan her problem-solving process and to purposefully integrate and apply her mathematical content knowledge about Cauchy sequences in her planning process. This may be seen as a result of the combined focus on two resources, problem solving and mathematical strategic knowledge, in the simultaneous intervention, as the conjunction of the results regarding both resources appear to have shaped her solution. Yet, the first line of her proof attempt also highlights that this newly-acquired, structured approach can also constrain her solution process as she did not just pursue her first, probably so far usual approach to proving the statement. The new approaches (analyzing the task according to the given prompts for each resource) and information on the resources require a certain amount of reprocessing and have to be integrated into Leia's prior knowledge and solution process for proof tasks. As she belongs to the low attaining students, the negotiation process between her prior knowledge and approaches with the new information introduced in the intervention on how to use the resources to analyze tasks, how to include the given prompts, and how to make meaning of the according results when working on a proof task, may reasonably not be as pronounced as for stronger students. Further, her prior approaches to constructing mathematical proofs may have been prone to error and even though the negation process may inhibit her to some degree, the gains will likely overweight. In comparison, high attaining students, who may have solved the task anyhow, will not benefit to this extent and the inhibition is likely to overweight.

5.3.6 Discussion

Our intervention study examined instructional approaches to support the learning of the resources underlying mathematical argumentation and proof skills, also aiming at benefits for the overall skill. For this, a *one-by-one* approach focusing on each resource individually and a *simultaneous* approach, focusing on the resources concurrently, both which were inspired by the part-task / whole-task debate from instructional design, were compared.

Effects on the resources: The preliminary analyses of the results revealed that explicit training of the resources of mathematical argumentation and proof skills can lead to remarkable learning gains when focusing on individual resources. The longitudinal effect sizes between the parallelized pre- and posttest indicate high positive effects on most resources, especially for mathematical strategic knowledge and problem-solving heuristics. These high effects on resources may reflect that both topics are not explicitly focused during university instruction on mathematics so that initial learning gains are easy to achieve.

Focusing on the differences between the impact of both approaches on the resources, a MANCOVA did not reveal an overall significant difference for students' cognitive resources. Solely students' learning of mathematical strategic knowledge showed a significantly higher gain in the simultaneous condition. Although our assumption was that the one-by-one approach would be superior for the learning of the resources, this result appears reasonable: Mathematical strategic knowledge refers to knowledge about cues within mathematical tasks that lead to promising methods or concepts to tackle the tasks, and further refers to knowledge about strategies to solve these tasks (Weber, 2001). It therefore is related to creating a representation of the problem (problem space), identifying operators therein, and choosing an operator that may be useful to accomplish the task. The successful use of mathematical strategic knowledge therefore corresponds to a rather comprehensive view of tasks and is not only limited to certain steps within the task. Mathematical strategic knowledge further shows multiple connections to the other resources, as for example mathematical content knowledge is needed to create the problem space and identify the operators. Further, methodological knowledge is needed to identify what a goal state for the problem is supposed to entail. Accordingly, the simultaneous approach could be beneficial for this resource as implied by the data as it may emphasize and strengthen these relations to other resources.

Effects on mathematical argumentation and proof skills: An ANCOVA comparing the posttest results regarding students' mathematical argumentation and proof skills did not show a significant difference between both approaches. Still, examining this result more closely by using a median-split revealed interesting effects: Whereas the initially stronger students could not benefit from the intervention, weaker students showed a positive development, having better learning gains for overall mathematical argumentation and proof skills. Here especially the students from the simultaneous approach could benefit, suggesting that the integration of the individual resources is of major importance and succeeds better when addressing the resources concurrently. This discrepancy between weaker and stronger students could be interpreted as an expertise reversal effect (Kalyuga, 2007; Kalyuga, Ayres, Chandler, & Sweller, 2003; Kalyuga & Renkl, 2010; Salden, Aleven, Schwonke, & Renkl, 2010), suggesting that stronger students need more effort to integrate the new information into their existing resources. Possibly, the findings may also be caused by a regression to the mean (see James, 1973).

5.3.7 Conclusions and Outlook

The current study highlights that acknowledging the nature of a complex cognitive skill, having multiple underlying resources, can inspire education and raises new questions for research. We present first evidence that supporting a complex cognitive skill can not only be achieved by focusing on the skill itself, but also with a *resource-based strategy*, explicitly addressing the resources underlying the complex cognitive skill.

The results show that in the case of mathematical argumentation and proof skills the one-by-one and the simultaneous approach can both be used to support students, at least regarding the resources. Here, both approaches yielded substantial learning gains, especially for mathematical strategic knowledge. Further, both approaches did not show large differences as implied by the part-task / whole task debate (Branch & Merrill, 2011; van Merriënboer & Kester, 2007), but are mostly comparable in learning gains.

Regarding the effects on students' mathematical argumentation and proof skills, further studies are needed. Quantitative studies with larger samples, qualitative studies focusing more intensely on the processes during the intervention as well as students' proof construction processes after the intervention, and studies examining the long-term effects of such an intervention will be able to further investigate the observed differences regarding prior attainment and examine whether also stronger students benefit from interventions targeting resources explicitly in the long run. The results of this study, therefore, have to be seen as the starting point of research regarding complex cognitive skills and their resources.

Further, including a control condition would be desirable to consolidate the results of this study. However, there is no generic candidate for this, as the resources are usually not taught in "regular" university mathematics courses. We would therefore rather propose to compare intervention approaches acknowledging the underlying resources with several other approaches, not explicitly taking the resources into account (e.g., Alcock & Simpson, 2002; Blanton et al., 2003; Heinze et al., 2008; Moore, 1994; Reiss et al., 2007; Samkoff & Weber, 2015; J. Selden & Selden, 1995). Outcomes could show whether acknowledging the underlying resources is beneficial for supporting students' learning. Here, special attention should be paid to the comparability of the interventions, for example by using academic learning time, time on task, or equivalent (Brodhagen & Gettinger, 2012) as a general measure.

Compared to other studies (e.g., Merrill, 2002) where the learning content was (almost) completely new to the learners, the current study is not concerned with the acquisition of a *new* complex cognitive skill. This may be seen as a partial limitation, but should be interpreted as the advantage of ecological validity. We examined real students in real learning settings, which are analogous to situations lecturers at university would face when trying to support their students' mathematical argumentation and proof skills. As these are the relevant situations, it is important to also examine these and not only lab settings.

Finally, the inclusion of further process measures is another important step for future research to understand the differences between both conditions better. Sociocultural and situated learning theories imply that students in the simultaneous condition should perceive the work on the tasks and accordingly their learning as more meaningful (e.g., Brown et al., 1989; Collins, Brown, & Newmann, 1989; Lave & Wenger, 1991), inducing positive emotions and task-values (see J. S. Eccles, 2005; Pekrun, 2006; Pekrun & Stephens, 2012), which may lead to productive behavior. Contrary, students may be overwhelmed and confused by this approach, possibly leading to less productive behavior. Thus, including measures for students' emotions and task-values would be valuable to observe the effects of the simultaneous condition on students better.

Our studies' main goal was to explore whether two different approaches (one-by-one and simultaneous) that were inspired by research from instructional design would yield different learning gains regarding a complex cognitive skill and its resources. The results of the study reveal that both approaches for supporting students' cognitive resources and their overall mathematical argumentation and proof skills, do not differ vastly in their effectiveness. Thus, findings therefore reveal that the tenet of the part-task / whole-task debate (J. R. Anderson et al., 1996; Branch & Merrill, 2011) that whole-task approaches are favorable in the context of complex cognitive skills cannot be transferred directly.

The study furthermore shows that including the resources into instruction supporting mathematical argumentation and proof skills is highly valuable for the learning of the resources and also has effects on students' mathematical argumentation and proof skills. Here, students with different prior attainment appear to benefit differently, and especially students with a low prior attainment benefit substantially from the simultaneous approach.

Although future research is needed, the study showed a new *resource-based approach* to develop powerful interventions that are not only limited to artificial settings with few students, but can be scaled up (e.g., Looi & Teh, 2015; Sternberg et al., 2006) and are suitable for whole classes or lectures.

6 Synthesis

Outline *We will first summarize and connect the central findings of our three studies and discuss them in the larger context of current research. Some limitations of the present studies will be outlined before completing this thesis by conclusions and an outlook on ongoing and future research based on the results of this project.*

We started the MIMAPS project to acquire a better understanding of the influence of individual, cognitive resources underlying mathematical argumentation and proof skills on first-year university students' performance and how such resources can be used to support students. Our central aims were to identify cognitive resources that were suggested by prior research to be underlying mathematical argumentation and proof skills and to empirically determine their relative impact on students' performance in handling mathematical argumentation and proof. These are of particular interest, as resources with a high, positive impact represent candidates for educational interventions to support the resources themselves as well as students' overall mathematical argumentation and proof skills.

To address these aims, we first created a novel framework for mathematical argumentation and proof skills (see section 4.1). It takes a comprehensive view focusing on three aspects that are prominent in research on mathematical argumentation and proof skills, the *resources* underlying mathematical argumentation and proof skills, the *processes* the skill is enacted in, and the *situations* they are required in. The framework is based on the work by Blömeke et al. (2015) on complex cognitive skills, using various frameworks from prior mathematics education research to adapt it to the context of mathematical argumentation and proof skills (see section 4.1).

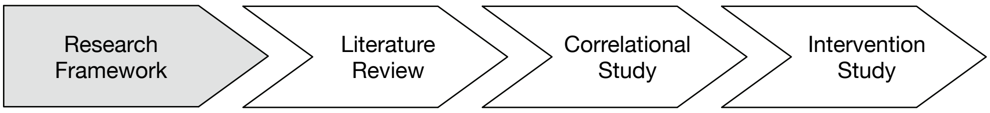


Figure 49. Illustration of the research framework and the three subsequent studies within our MIMAPS project.

Within our project, we conducted three consecutive studies spanning from theory to intervention (Figure 49). First, over 750 research reports were reviewed to obtain an overview of current emphases in research on mathematical argumentation and proof skills. Second, we examined first-year mathematics students and assessed six potential individual cognitive resources as well as students' performance in two situations requiring their mathematical argumentation and proof skills (proof construction and proof validation). Data were evaluated using Generalized Linear Mixed Models to determine the relative importance of each resource in either situation. Third, we compared two different instructional approaches for supporting the individual resources of mathematical argumentation and proof skills alongside the overall skill in a quasi-experimental intervention study.

Research was guided by the following questions that will also be used to structure the discussion:

RQ1 **Which resources, processes, and situations in the context of mathematical argumentation and proof skills are currently addressed by mathematics education research? Is there research focusing on the individual aspects of mathematical argumentation and proof skills in a comprehensive way? Which combinations of aspects have been examined?**

RQ2 **What is the relative influence of the potential individual resources underlying students' mathematical argumentation and proof skills on their performance in proof construction and validation? Can differences regarding the influence of domain-specific and domain-general resources be observed?**

RQ3 How do students' proof validation skills relate to their proof construction skills? Can their proof validation skills add to the explanation of their proof construction skills beyond the included individual resources?

RQ4 What are the differences between two instructional approaches that aim at supporting the resources individually one-by-one or simultaneously in terms of students' acquisition of individual resources as well as overall mathematical argumentation and proof skills?

6.1 Discussion of Central Findings

6.1.1 The Scope of Prior Research

Our initial systematic, descriptive literature review revealed that 20% of the 782 reviewed research reports focused on mathematical argumentation and proof, thus supporting the impression that the topic is central in mathematics education research (e.g., Hanna, 1991). The findings also show that research addresses all three central aspects of our research framework (resources, processes, and situations) on mathematical argumentation and proof skills, as well as all sub-aspects of these. However, large discrepancies regarding their importance in research can be seen. Currently, research focusing on the situation of *proof construction*, the process of *evidence generation*, and the resource *mathematical content knowledge* clearly dominates, replicating findings by Mejía-Ramos and Inglis (2009a, 2009b). Moreover, most studies take a narrow perspective on mathematical argumentation and proof skills. For example, only every fifth study explicitly considered more than one resource of mathematical argumentation and proof skills, and only 7% of the studies involved multiple situations of mathematical argumentation and proof skills. Accordingly, such studies help to better understand *individual aspects* but are very limited in their contribution to a *comprehensive view* of mathematical argumentation and proof skills. Furthermore, such studies are intrinsically incapable to compare the importance of the individual resources or to differentiate between the situations mathematical argumentation and proof skills are used in, as neither sufficient resources nor multiple situations are included.

These results of our literature review are underlined by the finding that many combinations of resources, processes, and situations relevant in the context of mathematical argumentation and proof skills were not addressed at all. This can be interpreted as evidence that these combinations are already well explored and that no further research is needed. However, we doubt this as for example there is little empirical knowledge regarding the influence of mathematical strategic knowledge (Weber, 2001) on proof construction and proof validation so far, but it proved to be an important predictive resource for students' performance in our correlational study (see section 5.2). Another possible interpretation is that these combinations of sub-aspects are examined but yield small effects or insights, so that this research is not reported due to publication bias (see Kühberger et al., 2014). Finally, the combinations that are currently not in the focus may just be deemed to be of limited interest and are therefore not examined.

Although there may be good reasons for *some* individual research gaps, our analyses of research on the combinations of resources, processes, and situations revealed that there are currently *many* combinations that appear not to be examined. Thus, there remain many open questions, for example regarding the impact of mathematical strategic knowledge on the different processes during proof construction such as hypothesis generation. Furthermore, not only our review revealed that the connections between the resources of mathematical argumentation and proof skills and students' performance in various situations appear to be under-researched so far, but also the results of the correlational study raise several new questions, for example regarding the varying influence of domain-general problem-solving skills depending on the availability of domain-specific heuristics. Thus, a systematic approach to the different connections between the aspects would be valuable.

6.1.2 The Resources Underlying Mathematical Argumentation and Proof Skills in Proof Construction and Proof Validation

6.1.2.1 Empirical Evidence for the Relevance of Individual Cognitive Resources

To examine the impact of resources underlying mathematical argumentation and proof skills, we assessed six cognitive resources as well as students' performance in proof validation and proof construction and used Generalized Linear Mixed Models to analyze the influence of the resources on either performance. The empirical findings showed that students' individual cognitive resources can indeed be used to model and predict their performance in proof construction and proof validation. Results thereby support our initial hypotheses and research framework regarding the status of mathematical argumentation and proof skills as a complex cognitive skill. Further, prior qualitative (e.g., Schoenfeld, 1985) and quantitative findings (e.g., Chinnappan et al., 2012; Ufer et al., 2008) regarding the necessity of underlying resources are verified and extended. In particular, we included a higher number of resources that were suggested as important for mathematical argumentation and proof skills, and included proof validation as another important situation requiring students' mathematical argumentation and proof skills.

Although results emphasize the importance of underlying resources, prior findings regarding the influence of the individual resources were only partially replicated. Our data verifies the influence of mathematical content knowledge on proof construction found in prior studies (Chinnappan et al., 2012; Ufer et al., 2008) but not the high influence of problem-solving skills. The latter may be due to the sample size and according restrictions to power in this study, yet still implying that the effect of problem-solving skills was lower as compared to both prior studies. Another potential cause for the differing results is the influence of the conceptualization and operationalization of the individual resources on their measured impact. Although each resource within our framework was guided by prior research and thus some coherence can be assumed, there is a certain ambiguity in prior research regarding some resources.

First, *mathematical strategic knowledge* introduced by Weber (2001) has to be better integrated into current frameworks and its relations to other constructs, for example domain-general problem-solving skills, conceptual and procedural knowledge, as well as other constructs that are mentioned in the context of mathematical strategic knowledge, have to be examined and clarified. Weber and Alcock (2004) also mentioned the selection of a more promising proof production style (syntactic, semantic) to approach a given task (see section 3.4.1) as another facet of mathematical strategic knowledge, which yet has to be properly integrated into the construct. Our project added to the conceptualization of mathematical strategic knowledge by giving the, to our knowledge, first quantitative operationalization and empirically showing (see Table 5) that it is mostly independent of the other facets as it shows no significant correlation to problem-solving skills.

Second, there are varying conceptualizations of problem-solving skills, some including metacognition as a part of problem-solving skills (e.g., as *control* or *monitoring*), others conceptualizing it as a resource relevant for problem solving, yet in principle more general (see Carlson & Bloom, 2005; Gok, 2010; Mayer, 1998; Schoenfeld, 1985; Schoenfeld, 1992). Here, a clearer separation and a common operationalization would be desirable, too.

Besides these ambiguities, our studies have clearly shown that it is possible to measure multiple resources of mathematical argumentation and proof skills reliably and mostly independently, as demonstrated with the six resources in the correlational study. This is an important prerequisite for further research on the resources underlying mathematical argumentation and proof skills and thus represents an advancement for this line of research.

6.1.2.2 Domain-specificity of Mathematical Argumentation and Proof Skills

The findings of this project, especially those from the GLMM analysis within the correlational study, underline that handling mathematical argumentation and proof is a knowledge intensive activity that largely depends on domain-specific knowledge. For proof construction, all three content- and domain-specific resources included in the study showed a significant impact, and for proof validation, conceptual and mathematical strategic knowledge were the most important resources. Prior research had already revealed the high impact of mathematical content knowledge for proof construction (Chinnappan et al., 2012; Ufer et al., 2008), but the impact of mathematical strategic knowledge is a major new result, which is in line with according qualitative research results from mathematics education (e.g., Reiss & Heinze, 2004; Weber, 2001) as well as views regarding scientific reasoning and argumentation in general that underline the need for strategic and meta-strategic knowledge (e.g., Sodian & Bullock, 2008, p. 432). Moreover, the analysis of our intervention study shows that mathematical strategic knowledge can be effectively trained and especially benefits from an integrated view on the resources of mathematical argumentation and proof skills. We therefore emphasize mathematical strategic knowledge as an important resource of mathematical argumentation and proof skills, which so far is under-researched, and (to our experience) mostly not explicitly covered in mathematics education at university, but can be fostered effectively.

6.1.2.3 The Importance to Differentiate Between Situations

So far, research on mathematical argumentation and proof skills in general and the underlying resources in particular, was mainly concerned with proof construction (e.g., Chinnappan et al., 2012; Mejía-Ramos & Inglis, 2009a; Ufer et al., 2008). This is reflected by the various frameworks for the underlying resources (see section 3.2) and is also empirically underlined by our descriptive literature review (see section 5.1; Figure 33, left). Accordingly, little is known about the impact of students' individual resources in situations other than proof construction.

To address this research gap, we used Generalized Linear Mixed Models (see Bolker et al., 2009; Zuur et al., 2009) to examine the influence of six individual cognitive resources in the situations of proof construction and proof validation. Results reveal that conceptual, procedural, and mathematical strategic knowledge are the three most important resources for proof construction, whereas proof validation relies mostly on conceptual knowledge and mathematical strategic knowledge. Accordingly, one major difference between both situations is the low impact of procedural knowledge on proof validation whereas it appears to be the most important resource for proof construction. Further, proof validation appears to rely more on metacognitive awareness and problem-solving skills than proof construction (see Table 7). However, as (our local) students are not used to proof validation tasks, we assume that students had no prior experience on how to address these tasks and possibly possessed no domain-specific strategies to approach these tasks. Likely, students therefore relied on general problem-solving heuristics, thereby leading to the observed influence of problem-solving skills. In contrast, students can be assumed to have already worked on various proof construction tasks, thereby possibly acquiring domain-specific strategies to approach these tasks, leading to the low impact of problem-solving skills on proof construction.

Overall, results reveal that although both situations show a clear emphasis of concept- and domain-specific resources, both also differ in the impact of the underlying resources. Thus, results suggest differences also for the mental processes applied in these situations. Future research therefore needs to distinguish between different types of situations and demands in the context of mathematical argumentation and proof. Also, more emphasis should be devoted to proof reading and proof presentation as these do not only show substantial differences to proof construction, but are also suggested to be especially important in learning settings (Mejía-Ramos & Inglis, 2009a).

6.1.2.4 The (Lacking) Influence of Problem-solving Skills

As studies by both Ufer et al. (2008) and Chinnappan et al. (2012) had shown an impact of problem-solving skills on students' performance in proof construction, the lack of impact on both proof validation and proof construction in this project is an unexpected finding. Potential reasons contributing to the observed lack of impact of problem-solving skills are outlined below, each opening new directions for future research.

6.1.2.4.1 Domain-specific *vs*. Domain-general Strategies

One key reason for the difference between our and prior studies may be related to the samples of the studies and the development, selection, and interplay of domain-specific and domain-general strategies in the context of mathematical argumentation and proof skills, which is up to now not satisfactorily understood. Both prior studies found a significant impact of problem-solving skills (Chinnappan et al., 2012; Ufer et al., 2008) examining secondary school pupils, that is novices in comparison to the university students examined in our project. It is reasonable to assume that the pupils relied to a greater extent on *weak, domain-general problem-solving strategies*, whereas students may have used *stronger, domain-specific strategies* (see further Newell, 1980) that they acquired in their remaining secondary school education and their first semester at university. Thus, the shift in employed strategies towards domain-specific strategies likely resulted in the lower impact of problem-solving skills. Future studies regarding the availability of both kinds of strategies and their adaptive selection and use by students would be valuable (see also Siegler, 1989). Furthermore, longitudinal or quasi-longitudinal studies could provide insights into the development of the use and selection of domain-specific strategies (related to mathematical strategic knowledge) and domain-general strategies (problem-solving heuristics) depending on participants' amount of expertise and domain-specific knowledge.

6.1.2.4.2 Relation to the Trichotomy of Proof Construction

Considering the theoretical background regarding the processes and phases in the context of mathematical argumentation and proof skills (see section 3.3), another explanation for the lack of impact of problem-solving skills arises: The items employed in this project for proof construction have a medium to high empirical difficulty. Still, students' difficulties may not be related to the *solving the problem* phase of the trichotomy of proof construction as students already knew similar tasks. The difficulties may be rather related to both other phases, that is creating a deductive chain of arguments and correctly writing down the final proof. Accordingly, students' difficulties when handling the proof construction tasks may not be related to domain-general problem-solving skills, as these are mainly required in the first phase of the trichotomy. Accordingly, problem-solving skills may only explain a low amount of the variance in performance in these tasks as suggested by our data.

6.1.2.4.3 Semantic, Syntactic, and Procedural Proof Production

A third possible explanation contributing to the lack of impact of problem-solving skills is based on prior research regarding different types of proof production (Weber & Alcock, 2004, 2009). Weber and Alcock's (2004) exploratory study revealed that undergraduate students appear to rely mostly on syntactic proof production styles, whereas more advanced doctoral students and professional algebraists use semantic styles to guide their proof production. Accordingly, the students in our study, especially the weaker students, potentially did not use a semantic proof production approaches that rely on an understanding of the problem as well as informal representations and approaches, but rather pursued syntactic and procedural proof production approaches. In contrast, pupils in prior studies (Chinnappan et al., 2012; Ufer et al., 2008) worked on geometry proofs that were explicitly linked to informal, non-symbolic representations, therefore substantially requiring their problem-solving skills.

This explanation corresponds to the presumed expertise reversal effect observed in the intervention study: Given that undergraduate students predominantly apply syntactic and procedural proof production approaches (Weber & Alcock, 2004), an explicit training supporting the resources underlying mathematical argumentation and proof skills and the use of prompts that scaffold the use of a semantic proof production approach, may influence students leading to the observed longitudinal effects (see Table 10). For initially stronger students who had already mastered syntactic and procedural proof production to a certain extent, the new information and approach to handle proofs may have been hard to integrate into their prior knowledge and well-established approaches. In comparison, weaker students, who were not yet able to work profoundly with all three proof production styles, may have benefited from the intervention as it enabled them to use semantic proof production to a greater extent.

The reliance on syntactic and especially procedural proof production approaches is also a potential explanation for why students' conditional reasoning skills did not show a significant impact in our analyses. Potentially, students do not work on the level of inferences about mathematical concepts but rather use formal rules, and therefore no impact could be measured (see A. Selden & Selden, 2003; Weber & Alcock, 2005). This finding may further correspond to those by Inglis and Alcock (2012) that novices employ less checking of between-the-line arguments and checking of warrants when reading proofs. Perhaps, the construct of conditional reasoning skills may be too narrow in the context of mathematical argumentation and proof and should be replaced by a more general conception of reasoning skills, for example including further aspects of first-order logic.

6.1.2.4.4 Differences in the Operationalization

Finally, the differences regarding the influence of problem-solving skills may also be due to differences in the items employed to assess students' problem-solving skills and the coding used to analyze students' answers, as these differ between both prior and our study:

In their study, Chinnappan et al. (2012) assessed problem-solving skills using items that were largely unrelated to mathematics[25], for example asking students how to find out food preferences of peers for a party (Chinnappan et al., 2012, p. 874). This ensured the domain-generality of the items and clearly separated them from the employed geometry proof items. However, Chinnappan et al. used the same coding procedure for both problem-solving items and the geometry proof items, which is based on the work by Senk (1989) on *geometry proof problems*.

Ufer et al. (2008) on the other hand used items for mathematics-related problem-solving skills based on Lin (2005), which are somewhat closer to mathematics. These involve, for example, basic geometrical properties such as symmetry but not the geometrical content or procedures that were required for the proof tasks. The coding of students' problem-solving tasks was then based on the identification and processing of central ideas within the tasks.

[25] This judgement assumes that the published example item, which at most required basic arithmetic operations, is representative for all items.

Although approaches in both studies differ, both appear as equally reasonable, although each operationalization also affords a reason for the observed impact of problem-solving skills in their results: The identical coding in Chinnappan et al. (2012) and the more mathematics related problem-solving items in Ufer et al. (2008).
In our study, we used a triangulation of both prior approaches: Problem-solving skills were assessed using mostly mathematics unrelated items, whereas coding was based on the *central ideas strategy* used by Ufer et al. (2008). The again differing operationalization in our project may therefore contribute to the observed lack of impact of problem-solving skills.
Reviewing items of all three studies, differences regarding two main characteristics can be observed: First, some items are more closely related to mathematics, whereas others are mainly domain-general. Second, items appear to require different sorts of problem solving. Some items require a certain “insight” in order to be solved and can be partially related to a Gestalt Theory conception of problem solving (e.g., Duncker & Lees, 1945; Kaplan & Simon, 1990), whereas other tasks rather require analytical skills (see example item of Chinnappan et al.), and even others require the use of several problem-solving operators and heuristics.
Overall, none of these three approaches including conzeptualizations, operationalizations, and coding can be judged as correct or incorrect, yet they highlight issues in current research that still have to be overcome.

6.1.3 Proof Construction *vs*. Proof Validation

Our initial analysis of the relationship between students’ proof validation skills and proof construction skills revealed a significant correlation between both, which replicates prior findings (e.g., Ufer, Heinze, Kuntze, et al., 2009). However, in our correlational study the GLMM analysis that included all six resources and proof validation skills did not show an impact of proof validation skills on proof construction skills. Thus, proof validation could not explain additional variance of proof construction beyond that explained by the six resources when jointly analyzing them. Accordingly, the correlation observed between both appears to be an artifact caused by one or multiple resource(s) jointly underlying both skills. As our results show that the correlation between proof construction and proof validation can be explained by the six resources included in our framework, no evidence points towards another common underlying resource such as methodological knowledge, which has repeatedly been suggested by prior research (e.g., A. Selden & Selden, 2003; Ufer, Heinze, Kuntze, et al., 2009). This does not completely rule out that methodological knowledge is another joint underlying resource, because power might not have been sufficient to detect it, or because the six resources, especially the domain-specific ones, may be confounded with methodological knowledge. Here, further research is needed.
Further, it appears that there is at least no strong relation between proof validation skills and proof construction skills beyond the common resources. Thus, even though proof validation is a sub-process of proof construction, it apparently does not add to proof construction beyond the underlying resources.

6.1.4 Approaches to Support the Resources

To explore the educational implications of interpreting mathematical argumentation and proof skills as a complex cognitive skill we compared two instructional approaches to foster the underlying resources. Although the intervention study was rather short with only four 2-hour sessions, three out of the four included resources (content knowledge, methodological knowledge, and mathematical strategic knowledge) showed substantial learning gains in both conditions. In contrast, problem-solving skills did not benefit from the short intervention. This appears reasonable, because compared to the three domain-specific knowledge facets, problem solving is a very general skill that requires several underlying resources and their coordination and integration (e.g., Schoenfeld, 1985). Thus, this result reflects prior findings, which revealed that supporting problem-solving skills demands a decent amount of time (e.g., Schoenfeld, 1992), and therefore was to be expected.

The effects of the short intervention on students' performance in proof construction are mixed. In contrast to initially stronger students, the initially weaker students in the simultaneous condition appear to benefit substantially from the intervention. Potentially, this effect may be interpreted as an expertise reversal effect (see Kalyuga, 2007; Kalyuga et al., 2003; Salden et al., 2010) or as a regression to the mean (see James, 1973). The results may suggest that the intervention bore a high complexity for the stronger students as they needed time to integrate the input into their prior knowledge and behavior. Thus, longer interventions that are less dense, possibly including a delayed posttest, would be preferable. Further, a study with a larger number of participants could reveal more nuanced effects on students' mathematical argumentation and proof skills and the (interaction) effects regarding prior attainment and instructional approach. In addition, more qualitative process data, for example in the form of audio / video recordings, would be valuable to examine weaker and stronger students' processes when handling proofs, examine how these processes develop throughout the intervention, and evaluate their influence on learning gains.

Our findings show that both instructional approaches employed in the intervention study are of mostly comparable effectiveness with respect to students' learning gains on the resources and their overall mathematical argumentation and proof skills. The one-by-one approach, which focused on each of the resources individually, and the simultaneous approach, which focused on them at the same time, only resulted in a significantly different learning gain in mathematical strategic knowledge, which was higher in the simultaneous approach. Thus, our data suggest that the large advantages of a whole-task approach, a central tenet of the part-task / whole-task debate in instructional design (e.g., Branch & Merrill, 2011; Fontana, Mazzardo, Furtado, & Gallagher, 2009; Lim, 2006; Lim et al., 2009; van Merriënboer & Kester, 2007), cannot be directly transferred to the level of individual cognitive resources and overall mathematical argumentation and proof skills, at least not in that size. Accordingly, more research is needed to examine how the acquisition and support of the individual resources underlying a complex cognitive skill transfer to the overall skill. Furthermore, it would be valuable to compare the *resource-based* approaches used in our intervention to other approaches (e.g., Alcock & Simpson, 2002; Blanton et al., 2003; Heinze et al., 2008; Moore, 1994; Reiss et al., 2007; Samkoff & Weber, 2015; J. Selden & Selden, 1995) in order to examine if an approach explicitly focusing on resources underlying students' mathematical argumentation and proof skills is superior to other instructional approaches.

6.2 Limitations

Besides the results of our three studies and their connections, some limitations need to be mentioned. First, there are common methodological and practical challenges: Naturally, a literature review is prone to bias resulting from the inclusion and exclusion of different sources and from the publication bias (see further Kühberger et al., 2014) within these sources themselves. Therefore, we have chosen to use conference proceedings from a recognized, international conference for mathematics education research, an approach that appears less biased to us then selecting specific journals.

Both empirical studies of this project are subject to the limitations of field experiments and ecological validity, such as the inability to randomly assign students to intervention groups. Yet, we deliberately conducted the intervention study in an ecologically valid way to ensure that the examined approaches are feasible in real-life situations and could be scaled up to whole classes and lectures. As both empirical studies were part of a voluntary course, they may further be prone to selection bias and re-sampling more students after the course was not possible. The project would certainly have benefited from larger sample sizes to increase the higher power of statistical hypothesis testing and to obtain more robust estimates of the coefficients for each resource (correlational study) and of effect sizes (intervention study). This limitation was partially overcome by using advanced statistical methods (Generalized Linear Mixed Models) and appropriate statistical parameters for small sample sizes in the correlational study. Finally, a larger number of students would have made it possible to include a control group for the intervention study.

Besides these rather generic limitations that apply to most empirical field research to a certain degree, we belief that the four more nuanced limitations outlined below may also be relevant for future research.

6.2.1 Conceptualization and Operationalization

An important basis for our studies was the selection of resources that were not only suggested to underlie mathematical argumentation and proof skills by prior research, but that also allowed a clear conceptualization, operationalization, and empirical separation of each other. Reviewing prior research (see also section 3.2) revealed a variety of studies and frameworks that were often using slightly different terms, conceptualizations, and operationalizations of the resources. In consequence, either specific conceptualizations had to be adopted, thereby disregarding others, or prior conceptualizations had to be triangulated by trying to distill common features – a problem also often encountered in reviews or meta-analyses (e.g., Cooper, Hedges, & Valentine, 2009; Cooper & Koenka, 2012). This lack of agreement on conceptualization and operationalization limits possibilities to compare results with prior research and to make general claims, a critique that was also mentioned for example for Hattie’s (2008) second-level meta-analysis (e.g., Snook, O’ Neill, Clark, O’ Neill, & Openshaw, 2009; Wecker, Vogel, & Hetmanek, 2017).

Accordingly, our results should be seen as dependent on the projects’ conceptualizations and operationalizations. As this limitation is unavoidable given the degree of agreement regarding the resources in prior research, we tried to be as transparent as possible regarding the conceptualization (see section 4.1.1), the operationalization, and the coding (see descriptions within the individual studies. Furthermore, we classified the resources using the frameworks by de Jong and Ferguson-Hessler (1996) and Chinnappan and Lawson (1996) to give a better description and overview. Although the results of this project underline that our approach was successful, a theory-based effort to create a unique framework for resources of mathematical argumentation and proof skills would be desirable for future research. For this, existing frameworks need to be systematically compared, analyzing similarities as well as differences. The resulting theory-based framework than needs to be aligned to already existing empirical data and empirically confirmed in subsequent studies to validate the included aspects and verify that ideally all central resources are included.

6.2.2 Coverage of Relevant Resources

The comprehensive research approach pursued throughout this project caused practical limitations in both experimental studies. We intended to measure several individual, cognitive resources as well as students' mathematical argumentation and proof skills in multiple situations in a way conforming to the performance criteria of empirical studies, for example ensuring reliability and validity (e.g., Bortz & Döring, 2006; Kantowitz, Roediger III, & Elmes, 2015). For this, the scales for the resources and for mathematical argumentation and proof skills in several situations had to each consist of sufficient items to span the conceptual breadth and allow a reliable measurement. However, testing time needed to be limited due to organizational reasons and, more significantly, due to mental fatigue, which can impact task performance (see Ackerman & Kanfer, 2009; Möckel, Beste, & Wascher, 2015; van der Linden, Frese, & Meijman, 2003) and accordingly may bias research findings. In consequence, the number of resources had to be limited to six in our correlational study and four in the intervention study, and students' time for each subscale was limited based on generous timings from a pilot study.

6.2.3 Disregarding the Processes

All three aspects (*resources*, *processes*, and *situations*) of the comprehensive research framework underlying our project were considered in the literature review, whereas both empirical studies "only" focused on several resources as well as two situations in the context of mathematical argumentation and proof skills. Although this still to our knowledge includes more resources and situations than any prior study, an assessment of the processes may have added significantly to the value of the empirical studies. Hence, future studies should try to incorporate an assessment of the processes into their research design, for example by video recording participants' actions. However, the inclusion entails further problems as participants' processes of interest are mental (see Blömeke et al., 2015) and measures based on observable characteristics can only be seen as a proxy measure for the former. Examining the example shown in section 2.4.1 reveals that many processes of interest are not directly accessible even though some are communicated orally or are enacted or embodied (e.g., Abrahmson & Lindgren, 2014; Alibali & Nathan, 2012), and therefore can be captured via video recordings. Furthermore, video recording students will fail to yield additional insights if students are working silently on given tasks and scales. Therefore, to capture the processes other methodological approaches are needed, for example using think-aloud-methods (see further Ericsson & Simon, 1998) or using collaborative settings and observing processes in dyads or groups (e.g., Kirsten, 2017; Nussbaum, 2008; Ottinger et al., 2017). Still, it also needs to be considered how these different approaches change students' behavior.

6.2.4 Resources Included in the Intervention Study

A limitation specific to our intervention study is related to the included resources. First, the number of resources had to be reduced to four in this study, as time was even more limited. However, there is not sufficient prior research on the teaching of resources underlying mathematical argumentation and proof skills to make a clear statement how this decision impacted the effects of our intervention. It may be a limitation, reducing the effects of the intervention, as not all resources from the working model are included, or be a benefit as complexity is reduced.

In consequence, we included *mathematical content knowledge*, *mathematical strategic knowledge*, *methodological knowledge*, and *problem-solving skills*. As results from the correlational study were not available at that time, the selection was based on prior existing

literature. As results of the correlational study revealed later, one of the four resources (problem-solving skills) was not validated as a resource underlying mathematical argumentation and proof skills. Further, methodological knowledge had not been included in the correlational study, so that no definite answer regarding its role can be given.
Overall, not all "resources" included in the intervention study were empirically validated as resources significantly predicting students' performance in proof construction in this study. Thus, results regarding the effect of the intervention on students' mathematical argumentation and proof skills might have been more conclusive if the study would have focused only on those resources that have been empirically underlined.

6.2.5 Structuring Mathematical Argumentation and Proof Skills Based on Several Resources, Processes, and Situations

One important assumption underlying this project and the research framework is that students' mathematical argumentation and proof skills can be analytically divided based on several resources, processes, and situations. This decomposition of mathematical argumentation and proof skills clearly is only a model. It is used to reduce the complexity of the phenomenon and to aid the understanding of mathematical argumentation and proof skills (Blömeke et al., 2015). It helps to structure a monolithic skill by creating smaller entities that are easier accessible, deemed to be key aspects for understanding the overall skill, and can help to explain the variance in students' performance. Although this approach has been taken in several other research areas with positive results (e.g., Sadler, 2013; Shavelson, 2010; Shulman, 1986), it can be questioned based on the phrase "the whole is more than the sum of its parts" (see Sadler, 2013, pp. 15-17). Furthermore, as our research framework (see section 4.1) and the correlational study (see section 5.2) only include a limited number of resources, which are assumed to be among the most important resources based on prior research, the model is likely not loss-free and cannot completely explain students' argumentation and proof skills. Yet, it is an empirical question how useful the approach is in describing overall mathematical argumentation and proof skills. Our project gives evidence that the framework can be used to structure research on argumentation and proof skills and that the approach appears to be adequate to model students' mathematical argumentation and proof skills. Further, based on the results from the intervention study, the approach also seems beneficial for instruction as learning effects on the resources were mostly positive. Still, whether the approach is also beneficial for the support of overall mathematical argumentation and proof skills remains to be verified in subsequent studies.

6.3 Outlook

6.3.1 Ensuring the Generalizability of the Results of the Project

Based on the limitations of the studies of our project, future replication studies as well as further analogue studies, for example including a longer intervention, are of high value. First, a replication of the experimental studies within this project with more participants will lead to a stronger empirical basis and give further evidence for the underlying framework and methodological approach. Given sufficient data, these studies also allow to use other than *linear* models to analyze the data in order to examine more nuanced effects of the resources, for example regarding very low or high availability of certain resources. Second, analogue studies in other mathematical content areas, for example, linear algebra, will show if the results of our studies, such as the relative influence of the resources, are (to some degree) concept-specific.

Third, a replication of the correlational study with more experienced students or experts will allow comparisons regarding the importance of the resources between these groups, potentially unveiling shifts in the importance of some resources with increasing mathematical expertise.
For this project, we decided to focus on first-year university students as research repeatedly highlighted their struggles, and we aimed to examine the resources underlying their mathematical argumentation and proof skills to identify those with a high relative impact, leading to the successful handling of argumentation and proof. This assured ecological validity and allows to "suggest learning trajectories that might be applicable for many other students as well" (Weber, 2009, p. 201). Focusing on more experienced students or experts in future studies will give insights into experts' use of resources and help to understand the influence on their performances in various situations. Still, the pedagogical value of such results are often limited (Weber, 2009), and research within the novice-expert paradigm has repeatedly shown that "it would be a mistake simply to expose novices to expert models and assume that the novices will learn effectively" (Donovan, Bransford, Pellegrino, & others, 1999). Accordingly, results could rather be used to compare the influence of the resources between the groups and to examine experts' mathematical argumentation and proof skills as the "goal-state" of university mathematics education, rather than trying to directly impose experts' behavior on students via instructional interventions.

6.3.2 Acknowledging Resources, Processes, and Situations in the Context of Mathematical Argumentation and Proof Skills

A key finding of our project is that first-year university students' mathematical argumentation and proof skills depend on several individual cognitive resources, and that their influence differs between situations. Thus, results underline the importance of resources and situations in the context of mathematical argumentation and proof skills. More generally, our MIMAPS project gave first evidence that the research framework based on the work by Blömeke et al. (2015) is valuable for planning, conducting, and analyzing research with a comprehensive view on mathematical argumentation and proof skills. Accordingly, further use of the framework is desirable. Primarily, the framework can be used to structure future studies with respect to the different aspects and sub-aspects of the framework, especially to safeguard that all relevant aspects of mathematical argumentation and proof skills that may bias research results are included, as for example underlying resources may lead to systematic differences in research findings. Moreover, the use of the framework would lead to an organized net of systematic research on mathematical argumentation and proof skills as each study could be positioned within the framework. Consequently, research would lead to an increasingly coherent and comprehensive picture of mathematical argumentation and proof skills and research gaps could be discovered more easily.
From an educational point of view, a major aim is to assure that students have the necessary resources at hand (see Schoenfeld, 2012b) that are required to effectively engage in their mathematics studies. This may be especially important at the transition from secondary to university education, as students enter university with different prior education. Another way to purposefully use the resources for instruction is by utilizing them to structure proof construction processes and thereby offer students a scaffold for their work. Thus, a training similar to self-explanation trainings (Hodds et al., 2014) may prove effective to support students' proof construction. Here, not only an analysis of the effects on students' performance, but also of the individual processes and how instruction regarding the resources may change and shape these, would be valuable.

Finally, the individual resources of mathematical argumentation and proof skills that showed a substantial impact within this project are mainly content- and domain-specific. Accordingly, researchers as well as lecturers may want to focus on such resources when working on approaches to support students. Here, the project provides evidence that this can be done effectively, at least on the level of the resources themselves and partially also for the overall skill. Yet, what the exact conditions for the effectiveness of interventions focusing on the resources of mathematical argumentation and proof skills are, was not answered definitely. Judging from the positive effects of the simultaneous approach, which focuses on various resources at the same time, this type of resource-based intervention seems to be particularly beneficial for weaker students. Here, more research, especially including longer intervention studies, which allow students to better process the input and annex it to their prior knowledge, is needed.

6.3.3 The Interplay of Proof Construction and Proof Validation

Within this study, we emphasized that there are different situations that require mathematical argumentation and proof skills. However, handling argumentation and proof in these situations can also be educationally used to support students' mathematical argumentation and proof skills as well as the individual resources. This can be strategically used in university teaching contexts. Besides comprehending proofs that are presented in lectures, students have so far been mainly asked to construct proofs for their lectures and seminars at university. In contrast, proof validation has been only implicitly included. However, research suggests that proof validation can inform proof construction (Pfeiffer, 2009a, 2011), can provide rich learning opportunities, and enables novice students to participate in mathematical practice (Pfeiffer & Quinlan, 2015). Our findings add to this by showing that students' performance in both situations partially depends on the same resources (i.e., conceptual mathematical knowledge, mathematical strategic knowledge). Thus, by handling mathematical argumentation and proof tasks in one situation, students indirectly also work on the underlying resources and thereby may reflect, elaborate, and train these resources, possibly leading to learning gains for the individual resources and their integration. At least the individual learning gains regarding each resource may likely transfer also to other situations involving mathematical argumentation and proof tasks which require the same resources. Thus, handling mathematical argumentation and proof tasks in one situations may be beneficial for other situations, too (Figure 50).

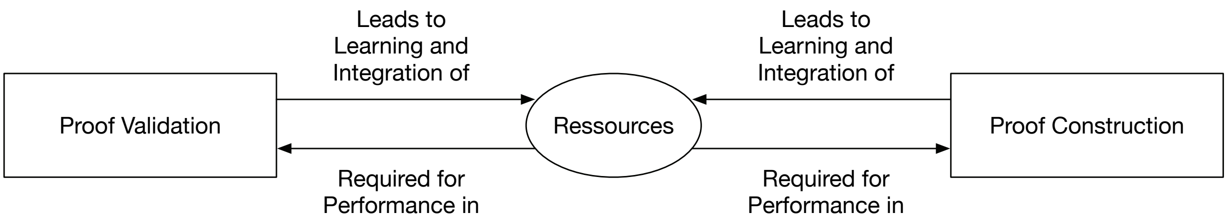


Figure 50. Reciprocal connection between proof validation and proof construction via the underlying resources.

Furthermore, our findings as well as prior research (e.g., Healy & Hoyles, 1998) suggest that students' have less difficulties with proof validation than with proof construction. Accordingly, proof validation tasks may render (the concept of) proof easier accessible for students when starting to handle proofs at the university and may be especially helpful to train specific resources such as mathematical strategic knowledge or methodological knowledge. Therefore, a *proof validation alongside proof construction* or *proof validation before proof construction* strategy may be useful for supporting students in acquiring mathematical argumentation and proof skills. As empirical evidence for the effectiveness of such strategies is missing, we have already conducted a mixed-methods study to address this question and data are currently in analysis.

6.3.4 The Trichotomy of Proof Construction

Considering both the trichotomy of proof construction (see section 3.3.4) and the lacking impact of problem-solving skills on students' proof construction performance in our project, further research regarding the three phases of the trichotomy is required. Although it has been repeatedly shown that constructing mathematical proofs is difficult across ages (e.g., Healy & Hoyles, 2000; A. Selden & Selden, 2013; Weber, 2001), these difficulties, more exactly the specific phase of the proof construction process that causes these difficulties, may depend on prior knowledge and problem-solving skills. As pointed out in section 6.1.2.4.2, first-year university students' difficulties with constructing proofs may not relate to the first step within the trichotomy (*solving the problem*), which mainly requires students' problem-solving skills, but may be due to problems with creating the deductive chain, formulating, and writing down the final proof. Here, insights in the relation between the three phases of the trichotomy, students' difficulties, and the need for problem-solving skills would be valuable. Results may be able to link the findings of our project with prior research (Chinnappan et al., 2012; Ufer et al., 2008) showing that problem-solving skills are an important resource for school pupils' mathematical argumentation and proof skills. The assumption that students' difficulties are related to a later phase in the trichotomy is also supported by positive effects of heuristic worked-out examples, which focus largely on problem solving, in school environments (e.g., Hilbert et al., 2008; Reiss et al., 2008), whereas effects at university appear to be lower (Kollar et al., 2014).
From an educational point of view, the trichotomy of proof construction may prove useful for instruction on proof construction, as it clearly separates three main phases of proof construction and could help students to understand, why their *solutions* for proof tasks are often not accepted as *proofs*. Here, methodological knowledge can also be explicitly picked up.

6.3.5 Giving Structure to the Resources

As pointed out in the limitations, the conceptualization, operationalization, and thereby the separation of the resources underlying mathematical argumentation and proof skills represent major obstacles (see section 6.1.2.4.4). Here, further clarification based, for example, on the frameworks by de Jong and Ferguson-Hessler (1996) or Chinnappan and Lawson (1996) would be valuable. Although our project has made a first step in this direction, future research integrating current theoretical frameworks is needed to identify, conceptualize, and empirically validate relevant resources of mathematical argumentation and proof skills. The resulting framework including conceptualizations, operationalization, and possibly exemplary scales of each resource would be an important basis to structure and compare future results.

6.4 Resume

The acquisition of mathematical argumentation and proof skills is one of the major learning goals of university mathematics programs, still mathematics students were repeatedly shown to struggle. Our MIMAPS project contributes to research on university students' argumentation and proof skills with several results and complements existing qualitative (e.g., Schoenfeld, 1985; Schoenfeld, 2012a; Weber, 2001) and quantitative (Chinnappan et al., 2012; Ufer et al., 2008) studies regarding the underlying resources. We give a descriptive overview of current research, highlighting three aspects (resources, processes, and situations) of mathematical argumentation and proof skills and point out currently under-researched areas. Further, empirical evidence on the relative importance of several cognitive resources underlying proof construction and proof validation is given, highlighting the importance of domain-specific knowledge and in particular of mathematical strategic knowledge for both situations. We also introduce two resource-based instructional approaches to support students in their acquisition of mathematical argumentation and proof skills, empirically quantifying and comparing the effects of both approaches on the resources and on overall mathematical argumentation and proof skills.

In all three consecutive studies ranging from literature review to intervention, the analytic approach to examine individual aspects instead of the overall complex cognitive skill has proven useful to explain the mechanisms underlying mathematical argumentation and proof skills and thus to obtain a better understanding. The overall framework introduced by our project, combining resources, processes, and situations in the context of mathematical argumentation and proof skills, can now be used as a fundament for future systematic research to construct an increasingly coherent picture of mathematical argumentation and proof skills as a complex cognitive skill.

7 References

Abel, C. F. (2003). Heuristics and Problem Solving. *New Directions for Teaching and Learning, 2003*(95), 53-58. doi:10.1002/tl.113

Aberdein, A. (2009). Mathematics and Argumentation. *Foundations of Science, 14*(1), 1-8. doi:10.1007/s10699-008-9158-3

Abrahmson, D., & Lindgren, R. (2014). Embodiment and embodied design. In R. K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences (2nd Edition)* (pp. 358-376). Cambridge, UK: Cambridge University Press.

Ackerman, P. L., & Kanfer, R. (2009). Test length and cognitive fatigue: an empirical examination of effects on performance and test-taker reactions. *Journal of Experimental Psychology: Applied, 15*(2), 163-181.

Alama, J., & Kahle, R. (2013). Checking proofs. In A. Aberdein & I. J. Dove (Eds.), *The argument of mathematics* (pp. 147-170). Dordrecht, Netherlands: Springer Science+Business Media

Alcock, L. (2009). e-Proofs: Student experience of online resources to aid understanding of mathematical proofs *Proceedings of the 12th Conference on Research in Undergraduate Mathematics Education. Special Interest Group of the Mathematical Association of America on Research in Undergraduate Mathematics Education*. Raleigh, NC.

Alcock, L. (2010). Mathematicians' perspectives on the teaching and learning of proof. In F. Hitt, D. Holton, & P. Thompson (Eds.), *Research in collegiate mathematics education VII* (pp. 63-91). Providence, RI: American Mathematical Society

Alcock, L., Attridge, N., Kenny, S., & Inglis, M. (2014). Achievement and behaviour in undergraduate mathematics: personality is a better predictor than gender. *Research in Mathematics Education, 16*(1), 1-17. doi:10.1080/14794802.2013.874094

Alcock, L., Bailey, T., Inglis, M., & Docherty, P. (2014). The ability to reject invalid logical inferences predicts proof comprehension and mathematics performance. In T. Fukawa-Connelly, G. Karakok, K. Kenne, & M. Zandieh (Eds.), *Proceedings of the 17th Annual Conference on Research in Undergraduate Mathematics Education* (pp. 376-383). Denver, CO.

Alcock, L., Hodds, M., Roy, S., & Inglis, M. (2015). Investigating and Improving Undergraduate Proof Comprehension. *Notices of the American Mathematical Society, 62*(7), 742-752.

Alcock, L., & Inglis, M. (2009). Representation systems and undergraduate proof production: A comment on Weber. *The Journal of Mathematical Behavior, 28*(4), 209-211. doi:10.1016/001j.jmathb.2009.10.001

Alcock, L., & Simpson, A. (2002). *Two components in Learning to Reason Using Definitions.* Paper presented at the 2nd International Conference on the Teaching of Mathematics (at the undergraduate level), Hersonissos, Greece.

Alcock, L., & Weber, K. (2005). Proof validation in real analysis: Inferring and checking warrants. *The Journal of Mathematical Behavior, 24*(2), 125-134. doi:10.1016/j.jmathb.2005.03.003

Alcock, L., & Weber, K. (2010). Referential and Syntactic approaches to proving: case studies from a transition-to-proof course. In F. Hitt, D. Holton, & P. Thompson (Eds.), *Research in Collegiate Mathematics Education. VII* (pp. 93-114). Providence, RI: American Mathematical Society.

Alcock, L., & Wilkinson, N. (2011). e-Proofs: Design of a Resource to Support Proof Comprehension in Mathematics. *Educational Designer, 1*(4). Retrieved from http://www.educationaldesigner.org/ed/volume1/issue4/article14/

Alexander, P. A., & Judy, J. E. (1988). The Interaction of Domain-Specific and Strategic Knowledge in Academic Performance. *Review of Educational Research, 58*(4), 375-404. doi:10.3102/00346543058004375

Alibali, M. W., & Nathan, M. J. (2012). Embodiment in Mathematics Teaching and Learning: Evidence From Learners' and Teachers' Gestures. *Journal of the Learning Sciences, 21*(2), 247-286. doi:10.1080/10508406.2011.611446

Alibert, D., & Thomas, M. (1991). Research on Mathematical Proof. In D. Tall (Ed.), *Advanced Mathematical Thinking* (Vol. 11, pp. 215-230). Dordrecht, Netherlands: Springer Science+Business Media.

Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological review, 89*(4), 369-406.

Anderson, J. R. (1983). *The Architecture of Cognition*. Cambridge, MA: Harvard University Press.

Anderson, J. R. (1993). *Rules of the Mind*. Mahwah, NJ: Lawrence Erlbaum Associates.

Anderson, J. R. (1996). ACT: A simple theory of complex cognition. *American Psychologist, 51*(4), 355-365. doi:10.1037/0003-066X.51.4.355

Anderson, J. R. (2002). Spanning seven orders of magnitude: a challenge for cognitive modeling. *Cognitive Science, 26*(1), 85-112. doi:10.1207/s15516709cog2601_3

Anderson, J. R., Reder, L. M., & Simon, H. A. (1996). Situated Learning and Education. *Educational Researcher, 25*(4), 5-11. doi:10.3102/0013189X025004005

Anderson, J. R., Reder, L. M., & Simon, H. A. (1997). Rejoinder: Situative versus cognitive perspectives: Form versus substance. *Educational Researcher, 26*(1), 18-21.

Anderson, L. W., Krathwohl, D. R., & Bloom, B. S. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. Boston, MA: Allyn & Bacon.

Anderson, R. C. (1968). Part-task versus Whole-task procedures for teaching a problem-solving skill to first graders. *Journal of Educational Psychology, 59*(3), 207-214. doi:10.1037/h0020050

Andriessen, J. E. B. (2009). Argumentation in Higher Education: Examples of Actual Practices with Argumentation Tools. In N. Muller Mirza & A.-N. Perret-Clermont (Eds.), *Argumentation and Education* (pp. 195-213). New York, NY: Springer US.

Attridge, N. (2013). *Advanced mathematics and deductive reasoning skills: Testing the theory of formal discipline (unpublished doctoral thesis)*. Loughborough University, UK.

Attridge, N., & Inglis, M. (2013). Advanced Mathematical Study and the Development of Conditional Reasoning Skills. *PLOS ONE, 8*(7). doi:10.1371/journal.pone.0069399

Balacheff, N. (1988). *Une étude des processus de preuve en mathématique chez des élèves de collège.* Univ. J. Fourier, Grenoble, France.

Balacheff, N. (1999). Is argumentation an obstacle? Invitation to a debate... Retrieved from http://www.lettredelapreuve.it/OldPreuve/Newsletter/990506Theme/990506ThemeUK.html

Bao, L., Cai, T., Koenig, K., Fang, K., Han, J., Wang, J., . . . Wu, N. (2009). Learning and Scientific Reasoning. *Science, 323*(5914), 586-587. doi:10.1126/science.1167740

Baroody, A. J., Feil, Y., & Johnson, A. R. (2007). An Alternative Reconceptualization of Procedural and Conceptual Knowledge. *Journal for Research in Mathematics Education, 38*(2), 115-131. doi:10.2307/30034952

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67*(1). doi:10.18637/jss.v067.i01

Beitlich, J. (2015). How do secondary school students make use of different representation formats in heuristic worked examples? An analysis of eye movements. In K. Beswick, T. Muir, & J. Wells (Eds.), *Proceedings of the 39th Conference of the International Group for the Psychology of Mathematics Education* (Vol. 2, pp. 97-104). Hobart, Australia: PME.

Beitlich, J., Obersteiner, A., Moll, G., Ruano, J. G. M., Pan, J., Reinhold, S., & Reiss, K. (2014). The Role Of Pictures In Reading Mathematical Proofs: An Eye Movement Study. In P. Liljedahl,

C. Nicol, S. Oesterle, & D. Allan (Eds.), *Proceedings of the 38th Conference of the International Group for the Psychology of Mathematics Education and the 36th Conference of the North American Chapter of the Psychology of Mathematics Education* (Vol. 2, pp. 121-128). Vancouver, Canada: PME.

Bell, A. W. (1976). A study of pupils' proof-explanations in mathematical situations. *Educational Studies in Mathematics, 7*(1-2), 23-40. doi:10.1007/BF00144356

Berland, L. K., & McNeill, K. L. (2010). A learning progression for scientific argumentation: Understanding student work and designing supportive instructional contexts. *Science Education, 94*(5), 765-793. doi:10.1002/sce.20402

Biehler, R., & Kempen, L. (2016). Conceptions of proof and proving in mathematics education: an analysis of the development of ideas in the didactics of mathematics. *Journal für Mathematik-Didaktik, 37*(1), 141-179. doi:10.1007/s13138-016-0097-1

Biggs, J. B. (1989). Approaches to the Enhancement of Tertiary Teaching. *Higher Education Research & Development, 8*(1), 7-25. doi:10.1080/0729436890080102

Black, P., & Wiliam, D. (2003). ‘In praise of educational research': formative assessment. *British Educational Research Journal, 29*(5), 623-637. doi:10.1080/0141192032000133721

Blanton, M. L., Stylianou, D. A., & David, M. (2003). The nature of scaffolding in undergraduates transition to formal proof. In N. Pateman, B. Daugherty, & J. Zilliox (Eds.), *Proceedings of the 27th Conference of the International Group for the Psychology of Mathematics Education* (Vol. 2, pp. 113-120). Honolulu, HI: PME.

Bleiler, S. K., Thompson, D. R., & Krajcevski, M. (2014). Providing written feedback on students' mathematical arguments: proof validations of prospective secondary mathematics teachers. *Journal of Mathematics Teacher Education, 17*(2), 105-127. doi:10.1007/s10857-013-9248-1

Blömeke, S., Gustafsson, J.-E., & Shavelson, R. J. (2015). Beyond Dichotomies. *Zeitschrift für Psychologie, 223*(1), 3-13.

Blum, W., Drüke-Noe, C., Hartung, R., & Köller, O. (2012). *Bildungsstandards Mathematik: konkret*: Humboldt-Universität zu Berlin, Institut zur Qualitätsentwicklung im Bildungswesen.

Boero, P. (1999, 20.11.2013). Argumentation and mathematical proof: A complex, productive, unavoidable relationship in mathematics and mathematics education. Retrieved from http://www.lettredelapreuve.org/OldPreuve/Newsletter/990708Theme/990708ThemeUK.html

Boero, P., Garuti, R., Lemut, E., & Mariotti, M. A. (1996). Challenging the traditional school approach to theorems: A hypothesis about the cognitive unity of theorems. In L. Puig (Ed.), *Proceedings of the 20th conference of the international group for the psychology of mathematics education* (Vol. 2, pp. 2-113). Valencia, Spain: PME.

Boero, P., Garuti, R., & Mariotti, M. A. (1996). Some dynamic mental processes underlying producing and proving conjectures. In L. Puig (Ed.), *Proceedings of the 20th conference of the international group for the psychology of mathematics education* (Vol. 2, pp. 121-128). Valencia, Spain: PME.

Bolker, B. M., Brooks, M. E., Clark, C. J., Geange, S. W., Poulsen, J. R., Stevens, M. H. H., & White, J.-S. S. (2009). Generalized linear mixed models: a practical guide for ecology and evolution. *Trends in ecology & evolution, 24*(3), 127-135.

Bortz, J., & Döring, N. (2006). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler, 4. überarb. Aufl.* Heidelberg, Deutschland: Springer.

Branch, R., & Merrill, M. D. (2011). Characteristics of instructional design models. In R. A. Reiser & J. V. Dempsey (Eds.), *Trends and issues in instructional design and technology* (3rd ed.). Upper Saddle River, NJ: Merrill-Prentice Hall.

Bransford, J., Sherwood, R., Vye, N., & Rieser, J. (1986). Teaching thinking and problem solving: Research foundations. *American Psychologist, 41*(10), 1078.

Brodhagen, E. M., & Gettinger, M. (2012). Academic Learning Time. In N. M. Seel (Ed.), *Encyclopedia of the Sciences of Learning* (pp. 33-36). Boston, MA: Springer US.

Brown, J. S., Collins, A., & Duguid, P. (1989). Situated Cognition and the Culture of Learning. *Educational Researcher, 18*(1), 32-42.

Bruner, J. S., Goodnow, J. J., & Austin, G. A. (1956). *A Study of Thinking*. Hoboken, NJ: John Wiley & Sons.

Brunner, E. (2013). *Innermathematisches Beweisen und Argumentieren in der Sekundarstufe I: mögliche Erklärungen für systematische Bearbeitungsunterschiede und leistungsförderliche Aspekte*. Münster, Germany: Waxmann Verlag.

Buchbinder, O., & Zaslavsky, O. (2013). Inconsistencies In Students' Understanding Of Proof And Refutation Of Mathematical Statements. In A. M. Lindmeier & A. Heinze (Eds.), *Proceedings of the 37th Conference of the International Group for the Psychology of Mathematics Education* (Vol. 2, pp. 129-136). Kiel, Germany: PME.

Byrne, O. (2010). The first six books of The Elements of Euclid in which coloured diagrams and symbols are used instead of letters for the greater ease of learners. Köln, Germany: (Reprinted by) Taschen.

Carlson, M., & Bloom, I. (2005). The Cyclic Nature of Problem Solving: An Emergent Multidimensional Problem-Solving Framework. *Educational Studies in Mathematics, 58*(1), 45-75. doi:10.1007/s10649-005-0808-x

Carrascal, B. n. (2015). Proofs, Mathematical Practice and Argumentation. *Argumentation, 29*(3), 305-324. doi:10.1007/s10503-014-9344-0

Chen, C.-H., & Chiu, C.-H. (2015). Collaboration Scripts for Enhancing Metacognitive Self-regulation and Mathematics Literacy. *International Journal of Science and Mathematics Education, 14*(2), 263-280. doi:10.1007/s10763-015-9681-y

Chi, M. T. H., Bassok, M., Lewis, M. W., Reimann, P., & Glaser, R. (1989). Self-Explanations: How Students Study and Use Examples in Learning to Solve Problems. *Cognitive Science, 13*(2), 145-182. doi:10.1207/s15516709cog1302_1

Chinn, C. A., & Clark, D. B. (2013). Learning through collaborative argumentation. In C. E. Hmelo-Silver, C. A. Chinn, C. K. K. Chan, & A. M. O'Donnell (Eds.), *International Handbook of Collaborative Learning* (pp. 314-332). New York, NY: Taylor & Francis.

Chinnappan, M., Ekanayake, M. B., & Brown, C. (2012). Knowledge use in the construction of geometry proof by sri lankan students. *International Journal of Science and Mathematics Education, 10*(4), 865-887. doi:10.1007/s10763-011-9298-8

Chinnappan, M., & Lawson, M. J. (1996). The effects of training in the use of executive strategies in geometry problem solving. *Learning and Instruction, 6*(1), 1-17. doi:10.1016/S0959-4752(96)80001-6

Christensen, R. H. B. (2015). ordinal: Regression Models for Ordinal Data.

Cilli-Turner, E. (2013). Effects of Collaborative Revision on Undergraduate Students' Proof Validation Skills. In M. Martinez & A. Castro Superfine (Eds.), *Proceedings of the 35th annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education* (pp. 272-275). Chicago, IL: University of Illinois at Chicago.

Cirillo, M., Kosko, K. W., Newton, J., Staples, M., & Weber, K. (2015). *Conceptions and Consequences of What We Call Argumentation, Justification, and Proof*. Paper presented at the 37th Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education, East Lansing, MI.

Cirillo, M., Kosko, K. W., Newton, J., Staples, M., Weber, K., Bieda, K., . . . Strachota, S. (2016). *Conceptions and Consequences of What We Call Argumentation, Justification, and Proof*. Retrieved from https://doi.org/10.13140/RG.2.1.5132.0727

Clark, K., James, A., & Montelle, C. (2014). "We definitely wouldn't be able to solve it all by ourselves, but together…'': group synergy in tertiary students': problem-solving practices. *Research in Mathematics Education, 16*(3), 306-323. doi:10.1080/14794802.2014.950687

Clark, M., & Lovric, M. (2008). Suggestion for a theoretical model for secondary-tertiary transition in mathematics. *Mathematics Education Research Journal, 20*(2), 25-37.

Clark, M., & Lovric, M. (2009). Understanding secondary-tertiary transition in mathematics. *International Journal of Mathematical Education in Science and Technology, 40*(6), 755-776.

Clark, R. E., & Estes, F. (1999). The Development of Authentic Educational Technologies. *Educational Technology, 39*(2), 5-16.

Collins, A., Brown, J. S., & Newmann, S. E. (1989). Cognitive Apprenticeship: Teaching the draft of reading, writing, and mathematics. In L. B. Resnick (Ed.), *Knowing, learning, and instruction: Essays in honor of Robert Glaser* (pp. 452-494). Hillsdale, NJ: Lawrence Erlbaum Associates.

Collins, A., Joseph, D., & Bielaczyc, K. (2004). Design research: Theoretical and methodological issues. *The Journal of the learning sciences, 13*(1), 15-42.

Common Core State Standards Initiative. (2010). *Common Core State Standards for Mathematics*. Washington, DC: National Governors Association Center for Best Practices and the Council of Chief State School Officers.

Conner, A., Singletary, L., Smith, R., Wagner, P., & Francisco, R. (2014). Teacher support for collective argumentation: A framework for examining how teachers support students' engagement in mathematical activities. *Educational Studies in Mathematics, 86*(3), 401-429. doi:10.1007/s10649-014-9532-8

Conradie, J., & Frith, J. (2000). Comprehension tests in mathematics. *Educational Studies in Mathematics, 42*(3), 225-235. doi:10.1023/A:1017502919000

Cooper, H., Hedges, L. V., & Valentine, J. C. (2009). *The handbook of research synthesis and meta-analysis*. New York, NY: Russell Sage Foundation.

Cooper, H., & Koenka, A. C. (2012). The overview of reviews: unique challenges and opportunities when research syntheses are the principal elements of new integrative scholarship. *American Psychologist, 67*(6), 446-462.

Corriveau, C., & Bednarz, N. (2017). The secondary-tertiary transition viewed as a change in mathematical cultures: an exploration concerning symbolism and its use. *Educational Studies in Mathematics, 95*(1), 1-19. doi:10.1007/s10649-016-9738-z

Csanadi, A., Kollar, I., & Fischer, F. (2016). Scientific Reasoning and Problem Solving in a Practical Domain: Are Two Heads Better Than One. In C. K. Looi, J. L. Polman, U. Cress, & P. Reimann (Eds.), *Transforming Learning, Empowering Learners: The International Conference of the Learning Sciences (ICLS) 2016* (Vol. 1, pp. 50-58). Singapore, Singapore: International Society of the Learning Sciences.

Csapó, B., & Szendrei, M. B. (2011). *Framework for diagnostic assessment of mathematics.* Budapest, Hungary Nemzeti Tankönyvkiadó.

Davis, P. J., & Hersh, R. (1981). *The mathematical experience*. New York, NY: Viking Penguin Inc.

Dawkins, P. C., & Weber, K. (2016). Values and norms of proof for mathematicians and students. *Educational Studies in Mathematics*. doi:10.1007/s10649-016-9740-5

De Corte, E., Verschaffel, L., & Op ’ t Eynde, P. (2000). Self-Regulation: A Characteristic and a Goal of Mathematics Education. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of Self-Regulation* (pp. 687-726). San Diego, CA: Academic Press.

De Guzmán, M., Hodgson, B. R., Robert, A., & Villani, V. (1998). Difficulties in the passage from secondary to tertiary education. In D. Biasius (Ed.), *Proceedings of the international Congress of Mathematicians* (Vol. 3, pp. 747-762). Berlin, Germany.

de Jong, T., & Ferguson-Hessler, M. G. M. (1996). Types and qualities of knowledge. *Educational Psychologist, 31*(2), 105-113. doi:10.1207/s15326985ep3102_2

de Villiers, M. (1990). The role and function of proof in mathematics. *Pythagoras, 24*, 17-24.

Devlin, K. (2003). When is a Proof? Retrieved from https://www.maa.org/external_archive/devlin/devlin_06_03.html

Di Martino, P., & Zan, R. (2011). Attitude towards mathematics: a bridge between beliefs and emotions. *Zentralblatt für Didaktik der Mathematik, 43*(4), 471-482. doi:10.1007/s11858-011-0309-6

Dieter, M., & Törner, G. (2012). Vier von fünf geben auf: Studienabbruch und Fachwechsel in der Mathematik. *Forschung und Lehre, 10*, 826-827.

Donovan, M. S., Bransford, J. D., Pellegrino, J. W., & others. (1999). *How people learn: Bridging research and practice*. Washington, DC: National Academy Press.

Dörfler, W., & McLone, R. R. (1986). Mathematics as a school subject. In B. Christiansen, A. G. Howson, & M. Otte (Eds.), *Perspectives on mathematics education* (pp. 49-97). Dordrecht, Netherlands: Springer Science+Business Media.

Dörner, D. (1979). *Problemlösen als Informationsverarbeitung*. Stuttgart, Germany: Kohlhammer

Douek, N. (2007). Some remarks about argumentation and proof. In P. Boero (Ed.), *Theorems in school: From history, epistemology and cognition to classroom practice* (pp. 163-181). Rotterdam, Netherlands: Sense Publishers.

Downs, M., & Mamona-Downs, J. (2005). The proof language as a regulator of rigor in proof, and its effect on student behavior. In M. Bosch (Ed.), *Proceedings of the Fourth Congress of the European Society for Research in Mathematics Education* (pp. 1748-1757). Sant Feliu de Guíxols, Spain.

Dreyfus, T., & Dvora, T. (2011). Unjustified assumptions in geometry. In B. Ubuz (Ed.), *Proceedings of the 35th Conference of the International Group for the Psychology of Mathematics Education* (Vol. 2, pp. 289-296). Ankara, Turkey: PME.

Dreyfus, T., Nardi, E., & Leikin, R. (2012). Forms of proof and proving in the classroom In G. Hanna & M. de Villiers (Eds.), *Proof and Proving in Mathematics Education: The 19th ICMI Study* (pp. 191-213). Dordrecht, Netherlands: Springer Science+Business Media.

Dunbar, K., & Klahr, D. (2012). Scientific Thinking and Reasoning. In K. J. Holyoak & R. G. Morrison (Eds.), *The Oxford Handbook of Thinking and Reasoning* (pp. 701-718). Oxford, UK: Oxford University Press.

Duncker, K., & Lees, L. S. (1945). On problem-solving. *Psychological monographs, 58*(5).

Duval, R. (1992). Argumenter, démontrer, expliquer: continuité ou rupture cognitive. *Petit x, 31*, 37-61.

Duval, R. (1995). *Sémiosis et pensée humaine: registres sémiotiques et apprentissages intellectuels*. Berne, Suisse: Peter Lang.

Eccles, D. W., & Feltovich, P. J. (2008). Implications of domain-general “psychological support skills'' for transfer of skill and acquisition of expertise. *Performance Improvement Quarterly, 21*(1), 43-60. doi:10.1002/piq.20014

Eccles, J. S. (2005). Subjective task value and the Eccles et al. model of achievement-related choices. In A. J. Elliot & C. S. Dweck (Eds.), *Handbook of competence and motivation* (pp. 105-121). New York, NY: The Guilford Press.

Epp, S. S. (2003). The role of logic in teaching proof. *American Mathematical Monthly, 110*(10), 886-899.

Erduran, S., & Jiménez-Aleixandre, M. P. (2008). *Argumentation in science education*. Dordrecht, Netherlands: Springer Science+Business Media

Ericsson, K. A., & Simon, H. A. (1998). How to Study Thinking in Everyday Life: Contrasting Think-Aloud Protocols With Descriptions and Explanations of Thinking. *Mind, Culture, and Activity, 5*(3), 178-186. doi:10.1207/s15327884mca0503_3

Evans, J. S. B. T., Clibbens, J., & Rood, B. (1995). Bias in Conditional Inference: Implications for Mental Models and Mental Logic. *The Quarterly Journal of Experimental Psychology Section A, 48*(3), 644-670. doi:10.1080/14640749508401409

Evans, J. S. B. T., Newstead, S. E., & Byrne, R. M. (1993). *Human reasoning: The psychology of deduction*. Hove, UK: Psychology Press.

Fallis, D. (2003). Intentional Gaps In Mathematical Proofs. *Synthese, 134*(1), 45-69. doi:10.1023/a:1022131513275

Felton, M., & Kuhn, D. (2001). The development of argumentive discourse skill. *Discourse Processes, 32*(2-3), 135-153.

Field, A. (2009). *Discovering statistics using SPSS*. New York, NY: Sage publications.

Fischer, F., Chinn, C. A., Engelmann, K., & Osborne, J. (Eds.). (in preparation). *Scientific Reasoning and Argumentation: Domain-Specific and Domain-General Aspects*: Taylor & Francis.

Fischer, F., Kollar, I., Ufer, S., Sodian, B., Hussmann, H., Pekrun, R., . . . Eberle, J. (2014). Scientific reasoning and argumentation: Advancing an interdisciplinary research agenda. *Frontline Learning Research, 4*, 28-45. doi:10.14786/flr.v2i2.96

Fischer, F., Wecker, C., Hetmanek, A., Osborne, J., Chinn, C. A., Duncan, R. G., . . . Sandoval, W. A. (2014). The Interplay of Domain-Specific and Domain-General Factors in Scientific Reasoning and Argumentation. In J. L. Polman, E. A. Kyza, D. K. O'Neill, I. Tabak, W. R. Penuel, A. S. Jurow, K. O'Connor, T. Lee, & L. D'Amico (Eds.), *International Conference of the Learning Sciences: Learning and Becoming in Practice* (Vol. 3, pp. 1189-1198). Boulder, CO: International Society of the Learning Sciences.

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American Psychologist, 34*(10), 906-911.

Fontana, F. E., Mazzardo, O., Furtado, J., Ovande, & Gallagher, J. D. (2009). Whole and part practice: a meta-analysis. *Perceptual and motor skills, 109*(2), 517-530. doi:10.2466/PMS.109.2.517-530

Fukawa-Connelly, T. (2012). A case study of one instructor's lecture-based teaching of proof in abstract algebra: making sense of her pedagogical moves. *Educational Studies in Mathematics, 81*(3), 325-345. doi:10.1007/s10649-012-9407-9

Fukawa-Connelly, T. (2014). Using Toulmin analysis to analyse an instructor's proof presentation in abstract algebra. *International Journal of Mathematical Education in Science and Technology, 45*(1), 75-88. doi:10.1080/0020739X.2013.790509

Furinghetti, F., & Morselli, F. (2009). Every unsuccessful problem solver is unsuccessful in his or her own way: affective and cognitive factors in proving. *Educational Studies in Mathematics, 70*(1), 71-90. doi:10.1007/s10649-008-9134-4

Furinghetti, F., & Pehkonen, E. (2002). Rethinking Characterizations of Beliefs. In G. C. Leder, E. Pehkonen, & G. Törner (Eds.), *Beliefs: A Hidden Variable in Mathematics Education?* (pp. 39-57). Dordrecht, Netherlands: Springer Netherlands.

Fuys, D., Geddes, D., & Welchman Tischler, R. (1984). English Translation of Selected Writings of Dina van Hiele-Geldof and Pierre M. van Hiele.

Gagné, R. M., & Merrill, M. D. (1990). Integrative goals for instructional design. *Educational Technology Research and Development, 38*(1), 23-30.

Geist, C., Löwe, B., & Van Kerkhove, B. (2010). Peer review and knowledge by testimony in mathematics. In B. Löwe & T. Müller (Eds.), *PhiMSAMP: Philosophy of mathematics: Sociological aspects and mathematical practice* (pp. 155-178). London, UK: College Publications.

Ghanem, C., Kollar, I., Fischer, F., Lawson, T. R., & Pankofer, S. (2016). How do social work novices and experts solve professional problems? A micro-analysis of epistemic activities and the use of evidence. *European Journal of Social Work*. doi:10.1080/13691457.2016.1255931

Giaquinto, M. (2005). Mathematical Activity. In P. Mancosu, K. F. Jørgensen, & S. A. Pedersen (Eds.), *Visualization, Explanation and Reasoning Styles in Mathematics* (Vol. 327, pp. 75-87). Dordrecht, Netherlands: Springer Netherlands.

Giere, R. N., Bickle, J., & Mauldin, R. (2005). *Understanding scientific reasoning* (5th ed.). Boston, MA: Cengage Learning.

Gigerenzer, G. (2008). Why Heuristics Work. *Perspectives on Psychological Science, 3*(1), 20-29. doi:10.1111/j.1745-6916.2008.00058.x

Gok, T. (2010). The General Assessment of Problem Solving Processes in Physics Education. *Eurasian Journal of Physics and Chemistry Education, 2*(2), 110-122.

Goldin, G. A. (1998). Representational systems, learning, and problem solving in mathematics. *The Journal of Mathematical Behavior, 17*(2), 137-165. doi:10.1016/S0364-0213(99)80056-1

Goodwin, C. (1994). Professional Vision. *American Anthropologist, 96*(3), 606-633. doi:10.1525/aa.1994.96.3.02a00100

Gravetter, F. J., & Wallnau, L. B. (2016). *Statistics for the behavioral sciences*. Boston, MA: Cengage Learning.

Greeno, J. G. (1997). On Claims That Answer the Wrong Questions. *Educational Researcher, 26*(1), 5-17. doi:10.3102/0013189X026001005

Greeno, J. G. (1998). The Situativity of Knowing, Learning, and Research. *American Psychologist, 53*(1), 5-26.

Greiff, S., Holt, D. V., & Funke, J. (2013). Perspectives on problem solving in educational assessment: Analytical, interactive, and collaborative problem solving. *Journal of Problem Solving, 5*(2), 71-91.

Greiff, S., Wüstenberg, S., Csapó, B. H., Demetriou, A., Hautamäki, J., Graesser, A. C., & Martin, R. (2014). Domain-general problem solving skills and education in the 21st century. *Educational Research Review, 13*, 74-83.

Griffin, T., Wiley, J., & Salas, C. (2013). Supporting Effective Self-Regulated Learning: The Critical Role of Monitoring. In R. Azevedo & V. Aleven (Eds.), *International Handbook of Metacognition and Learning Technologies* (Vol. 28, pp. 19-34). New York, NY: Springer US.

Grootenboer, P., & Marshman, M. (2016). The Affective Domain, Mathematics, and Mathematics Education *Mathematics, Affect and Learning: Middle School Students' Beliefs and Attitudes About Mathematics Education* (pp. 13-33). Singapore, Singapore: Springer Singapore.

Grünwald, N., Kossow, A., Sauerbier, G., & Klymchuk, S. (2004). Der Übergang von der Schul- zur Hochschulmathematik: Erfahrungen aus internationaler und deutscher Sicht. *Global Journal of Engineering Education, 8*(3), 283-294.

Gueudet, G. (2008). Investigating the secondary-tertiary transition. *Educational Studies in Mathematics, 67*(3), 237-254. doi:10.1007/s10649-007-9100-6

Hales, T. C. (2002). A computer verification of the Kepler conjecture. In T. D. Li (Ed.), *Proceedings of the ICM* (Vol. 3, pp. 795-804). Beijing, China.

Hales, T. C. (2008). Formal proof. *Notices of the AMS, 55*(11), 1370-1380.

Halpern, D. F. (1998). Teaching critical thinking for transfer across domains: Dispositions, skills, structure training, and metacognitive monitoring. *American Psychologist, 53*, 449-455.

Halpern, D. F. (2002). *Thought and knowledge: An introduction to critical thinking*. Mahwah, NJ: Lawrence Erlbaum Associates

Hanna, G. (1989). More than Formal Proof. *For the Learning of Mathematics, 9*(1), 20-23.

Hanna, G. (1990). Some pedagogical aspects of proof. *Interchange, 21*, 6-13.

Hanna, G. (1991). Mathematical Proof. In D. Tall (Ed.), *Advanced Mathematical Thinking* (Vol. 11, pp. 54-61). Dordrecht, Netherlands: Springer Netherlands.

Hanna, G. (2000). Proof, explanation and exploration: An overview. *Educational Studies in Mathematics, 44*(1-2), 5-23.

Hanna, G., & Jahnke, H. N. (1993). Proof and application. *Educational Studies in Mathematics, 24*(4), 421-438. doi:10.1007/BF01273374

Hanna, G., & Jahnke, H. N. (1996). Proof and proving. In A. Bishop, M. A. K. Clements, C. Keitel-Kreidt, J. Kilpatrick, & C. Laborde (Eds.), *International handbook of mathematics education* (pp. 877-908). Dordrecht, Netherlands: Springer Science+Business Media

Hannula, M. S. (2006). Affect in mathematical thinking and learning. In J. Maasz & W. Schloeglmann (Eds.), *New mathematics education research and practice* (pp. 209-232). Rotterdam, Netherlands: Sense Publishers.

Hannula, M. S., Di Martino, P., Pantziara, M., Zhang, Q., Morselli, F., Heyd-Metzuyanim, E., . . . Goldin, G. A. (2016). Attitudes, Beliefs, Motivation, and Identity in Mathematics Education *Attitudes, Beliefs, Motivation and Identity in Mathematics Education: An Overview of the Field and Future Directions* (pp. 1-35). Cham, Germany: Springer International Publishing.

Harel, G. (1999). Students' understanding of proofs: a historical analysis and implications for the teaching of geometry and linear algebra. *Linear Algebra and its Applications, 302-303*, 601-613. doi:10.1016/S0024-3795(99)00139-1

Harel, G., & Sowder, L. (1998). Students' proof schemes: results from exploratory studies. In A. H. Schoenfeld, J. Kaput, E. Dubinsky, & T. Dick (Eds.), *Research in collegiate mathematics education. III* (pp. 234-283). Providence, RI: American Mathematical Society.

Hattie, J. (2008). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. New York, NY: Routledge.

Healy, L., & Hoyles, C. (1998). *Technical Report on the Nationwide Survey. Justifiying and Proving in School Mathematics*. London, UK: Institute of Education, University of London.

Healy, L., & Hoyles, C. (2000). A Study of Proof Conceptions in Algebra. *Journal for Research in Mathematics Education, 21*(4), 396-428.

Heintz, B. (2000). *Die Innenwelt der Mathematik*. Wien, Austria: Springer.

Heinze, A. (2007). Problemlösen im mathematischen und außermathematischen Kontext. *Journal für Mathematik-Didaktik, 28*(1), 3-30. doi:10.1007/BF03339331

Heinze, A. (2010). Mathematicians' Individual Criteria for Accepting Theorems and Proofs: An Empirical Approach. In G. Hanna, H. N. Jahnke, & H. Pulte (Eds.), *Explanation and Proof in Mathematics: Philosophical and Educational Perspectives* (pp. 101-111). Boston, MA: Springer US.

Heinze, A., Cheng, Y.-H., Ufer, S., Lin, F.-L., & Reiss, K. (2008). Strategies to foster students' competencies in constructing multi-steps geometric proofs: teaching experiments in

Taiwan and Germany. *Zentralblatt für Didaktik der Mathematik, 40*(3), 443-453. doi:10.1007/s11858-008-0092-1

Heinze, A., & Kwak, J. Y. (2002). Informal prerequisites for informal proofs. *Zentralblatt für Didaktik der Mathematik, 34*(1), 9-16.

Heinze, A., & Reiss, K. (2003). Reasoning and Proof: Methodological Knowledge as a Component of Proof Competence. In M. A. Mariotti (Ed.), *Proceedings of the Third Conference of the European Society for Research in Mathematics Education* (Vol. 4, pp. 1-10). Bellaria, Italien.

Heinze, A., & Reiss, K. (2007). Reasoning and proof in the mathematics classroom. *Analysis, 27*(2-3), 333-357.

Heinze, A., & Reiss, K. (2009). Developing argumentation and proof competencies in the mathematics classroom. In D. A. Stylianou, M. L. Blanton, & E. J. Knuth (Eds.), *Teaching and Learning of Proof Across the Grades: A K-16 Perspective* (pp. 191-203). New York, NY: Routledge.

Heinze, A., Reiss, K., & Rudolph, F. (2005). Mathematics achievement and interest in mathematics from a differential perspective. *Zentralblatt für Didaktik der Mathematik, 37*(3), 212-220.

Hellmich, F., Hartmann, J., & Reiss, K. (2002). Bedingungen für das Argumentieren und Begründen im Geometrieunterricht. *Contributions to mathematics education 2002* (pp. 231-234). Hildesheim, Germany: Franzbecker.

Hembree, R. (1990). The Nature, Effects, and Relief of Mathematics Anxiety. *Journal for Research in Mathematics Education, 21*(1), 33-46. doi:10.2307/749455

Hemmi, K. (2008). Students' encounter with proof: the condition of transparency. *Zentralblatt für Didaktik der Mathematik, 40*(3), 413-426. doi:10.1007/s11858-008-0089-9

Hemmi, K. (2010). Three styles characterising mathematicians' pedagogical perspectives on proof. *Educational Studies in Mathematics, 75*(3), 271-291. doi:10.1007/s10649-010-9256-3

Herbst, P., Miyakawa, T., & Chazan, D. (2010). Revisiting the functions of proof in mathematics classrooms: A view from a theory of instructional exchanges.

Herppich, S., Praetorius, A.-K., Hetmanek, A., Glogger-Frey, I., Ufer, S., Leutner, D., . . . Südkamp, A. (2017). Ein Arbeitsmodell für die empirische Erforschung der diagnostischen Kompetenz von Lehrkräften. In A. Südkamp & A.-K. Praetorius (Eds.), *Diagnostische Kompetenz von Lehrkräften* (pp. 75-93). Münster, Germany: Waxmann.

Hersh, R. (1999). *What is Mathematics, really?* Oxford, UK: Oxford University Press.

Heublein, U., Richter, J., Schmelzer, R., & Sommer, D. (2012). Die Entwicklung der Studienabbruchquoten an den deutschen Hochschulen *Statistische Berechnungen auf der Basis des Absolventenjahrgangs 2010 (Forum Hochschule 3|2012)*. Hannover, Germany: DZHW.

Heublein, U., Richter, J., Schmelzer, R., & Sommer, D. (2014). Die Entwicklung der Studienabbruchquoten an den deutschen Hochschulen *Statistische Berechnungen auf der Basis des Absolventenjahrgangs 2012 (Forum Hochschule 4|2014)*. Hannover, Germany: DZHW.

Hiebert, J. (1986). *Conceptual and procedural knowledge: The case of mathematics*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Hiebert, J., & Lefevre, P. (1986). Conceptual and procedural knowledge in mathematics: An introductory analyis. In J. Hibert (Ed.), *Conceptual and procedural knowledge: The case of mathematics* (pp. 1-23). Hillsdale, NJ: Lawrence Erlbaum Associates

Hilbert, T. S., Renkl, A., Kessler, S., & Reiss, K. (2008). Learning to prove in geometry: Learning from heuristic examples and how it can be supported. *Learning and Instruction, 18*(1), 54-65. doi:10.1016/j.learninstruc.2006.10.008

Hodds, M., Alcock, L., & Inglis, M. (2014). Self-Explanation Training Improves Proof Comprehension. *Journal for Research in Mathematics Education, 45*(1), 62-101.

Hornikx, J., & Hahn, U. (2012). Reasoning and argumentation: Towards an integrated psychology of argumentation. *Thinking & Reasoning, 18*(3), 225-243. doi:10.1080/13546783.2012.674715

Iannone, P., & Inglis, M. (2010). Undergraduate students’ use of deductive arguments to solve ’prove that…' tasks. In M. Pytlak, T. Rowland, & E. Swoboda (Eds.), *Proceedings of the 7th Congress of the European Society for Research in Mathematics Education* (pp. 2112-2021). Rzeszów, Poland.

Inglis, M., & Aberdein, A. (2015). Beauty Is Not Simplicity: An Analysis of Mathematicians' Proof Appraisals. *Philosophia Mathematica, 23*(1), 87-109.

Inglis, M., & Alcock, L. (2012). Expert and novice approaches to reading mathematical proofs. *Journal for Research in Mathematics Education, 43*(4), 358-390.

Inglis, M., & Mejía-Ramos, J. P. (2009). On the Persuasiveness of Visual Arguments in Mathematics. *Foundations of Science, 14*(1-2), 97-110. doi:10.1007/s10699-008-9149-4

Inglis, M., Mejía-Ramos, J. P., Weber, K., & Alcock, L. (2013). On Mathematicians' Different Standards When Evaluating Elementary Proofs. *Topics in Cognitive Science, 5*(2), 270-282. doi:10.1111/tops.12019

Inglis, M., & Simpson, A. (2008). Conditional inference and advanced mathematical study. *Educational Studies in Mathematics, 67*(3), 187-204. doi:10.1007/s10649-007-9098-9

Inglis, M., & Simpson, A. (2009). Conditional inference and advanced mathematical study: further evidence. *Educational Studies in Mathematics, 72*(2), 185-198. doi:10.1007/s10649-009-9187-z

Inhelder, B., & Piaget, J. (1958). *The growth of logical thinking from childhood to adolescence: An essay on the construction of formal operational structure*. London, UK: Routledge & Kegan Pau.

Jacinto, H., & Carreira, S. (2013). Beyond-School Mathematical Problem Solving: A Case Of Students-With-Media. In A. M. Lindmeier & A. Heinze (Eds.), *Proceedings of the 37th Conference of the International Group for the Psychology of Mathematics Education* (Vol. 3, pp. 105-112). Kiel, Germany: PME.

Jahnke, H. N. (2007). Proofs and hypotheses. *Zentralblatt für Didaktik der Mathematik, 39*(1-2), 79-86.

Jahnke, H. N. (2010). The Conjoint Origin of Proof and Theoretical Physics. In G. Hanna, H. N. Jahnke, & H. Pulte (Eds.), *Explanation and Proof in Mathematics* (pp. 17-32). New York, NY: Springer US.

Jahnke, H. N., & Ufer, S. (2015). Argumentieren und Beweisen. In R. Bruder, L. Hefendehl-Hebeker, B. Schmidt-Thieme, & H.-G. Weigand (Eds.), *Handbuch der Mathematikdidaktik* (pp. 331-355). Berlin, Germany: Springer.

James, K. E. (1973). Regression toward the Mean in Uncontrolled Clinical Studies. *Biometrics, 29*(1), 121-130. doi:10.2307/2529681

Jaworski, B., Treffert-Thomas, S., & Bartsch, T. (2009). Characterising the teaching of university mathematics: A case of linear Algebra. In M. Tzekaki, M. Kaldrimidou, & H. Sakonidis (Eds.), *Proceedings of the 33rd Conference of the International Group for the Psychology of Mathematics Education* (Vol. 3, pp. 249-256). Thessaloniki, Greece.

Johnson-Laird, P. N. (1980). Mental models in cognitive science. *Cognitive science, 4*(1), 71-115.

Johnson-Laird, P. N. (2000). Thinking: Reasoning. In A. E. Kazdin (Ed.), *Encyclopedia of Psychology, Vol. 8* (pp. 75-79). Washington, D.C.: American Psychological Association.

Johnson-Laird, P. N., & Byrne, R. M. (2002). Conditionals: a theory of meaning, pragmatics, and inference. *Psychological review, 109*(4), 646-678.

Jones, K. (2000). The student experience of mathematical proof at university level. *International Journal of Mathematical Education in Science and Technology, 31*(1), 53-60. doi:10.1080/002073900287381

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: from eye fixations to comprehension. *Psychological review, 87*(4), 329-354.

Kalyuga, S. (2007). Expertise Reversal Effect and Its Implications for Learner-Tailored Instruction. *Educational Psychology Review, 19*(4), 509-539. doi:10.1007/s10648-007-9054-3

Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The Expertise Reversal Effect. *Educational Psychologist, 38*(1), 23-31. doi:10.1207/S15326985EP3801_4

Kalyuga, S., & Renkl, A. (2010). Expertise reversal effect and its instructional implications: introduction to the special issue. *Instructional Science, 38*(3), 209-215. doi:10.1007/s11251-009-9102-0

Kantowitz, B. H., Roediger III, H. L., & Elmes, D. G. (2015). *Research methods in psychology*. Stanford, CT: Cengage Learning.

Kapa, E. (2001). A Metacognitive Support during the Process of Problem Solving in a Computerized Environment. *Educational Studies in Mathematics, 47*(3), 317-336. doi:10.1023/A:1015124013119

Kaplan, C. A., & Simon, H. A. (1990). In search of insight. *Cognitive Psychology, 22*(3), 374-419. doi:https://doi.org/10.1016/0010-0285(90)90008-R

Kelly, C. A. (2006). Using manipulatives in mathematical problem solving: A performance-based analysis. *The Mathematics Enthusiast, 3*(2), 184-193.

Kempen, L. (2016). Beweisakzeptanz bei Studienanfängern: Eine empirische Untersuchung. In I. für Mathematik und Informatik Heidelberg (Ed.), *Beiträge zum Mathematikunterricht 2016*. Münster, Germany: WTM-Verlag.

Khine, M. S. (2012). *Perspectives on Scientific Argumentation: Theory, Practice and Research*. Dordrecht, Netherlands: Springer Netherlands.

Kirsten, K. (2017). Identifizierung von Phasen und Aktivitäten im Beweisprozess von Studienanfänger/innen. In Institut für Mathematik der Universität Potsdam (Ed.), *Beiträge zum Mathematikunterricht 2017*. Münster, Germany: WTM-Verlag.

Klahr, D., & Dunbar, K. (1988). Dual Space Search During Scientific Reasoning. *Cognitive Science, 12*(1), 1-48. doi:10.1207/s15516709cog1201_1

Klieme, E., & Leutner, D. (2006). Kompetenzmodelle zur Erfassung individueller Lernergebnisse und zur Bilanzierung von Bildungsprozessen: Beschreibung eines neu eingerichteten Schwerpunktprogramms der DFG. *Zeitschrift für Pädagogik, 52*(6), 876-903.

Kloosterman, P., Raymond, A. M., & Emenaker, C. (1996). Students' Beliefs about Mathematics: A Three-Year Study. *The Elementary School Journal, 97*(1), 39-56.

KMK. (2012). *Bildungsstandards im Fach Mathematik für die Allgemeine Hochschulreife. Beschluss vom 18.10.2012*. Bonn, Germany: (Beschlüsse der Kultusministerkonferenz).

Knipping, C. (2008). A method for revealing structures of argumentations in classroom proving processes. *Zentralblatt für Didaktik der Mathematik, 40*(3), 427-441. doi:10.1007/s11858-008-0095-y

Ko, Y.-Y., & Knuth, E. J. (2013). Validating proofs and counterexamples across content domains: Practices of importance for mathematics majors. *The Journal of Mathematical Behavior, 32*(1), 20-35.

Koedinger, K. R. (1998). Conjecturing and argumentation in high-school geometry students. In R. Lehrer & D. Chazan (Eds.), *Designing learning environments for developing understanding of geometry and space* (pp. 319-347). Mahwah,NJ: Lawrence Erlbaum Association.

Koeppen, K., Hartig, J., Klieme, E., & Leutner, D. (2008). Current Issues in Competence Modeling and Assessment. *Zeitschrift für Psychologie, 216*(2), 61-73.

Koichu, B., Berman, A., & Moore, M. (2007). Heuristic literacy development and its relation to mathematical achievements of middle school students. *Instructional Science, 35*(2), 99-139. doi:10.1007/s11251-006-9004-3

Koichu, B., & Leron, U. (2015). Proving as problem solving: The role of cognitive decoupling. *The Journal of Mathematical Behavior, 40*(Part B), 233-244. doi:10.1016/j.jmathb.2015.10.005

Kollar, I., Ufer, S., Reichersdorfer, E., Vogel, F., Fischer, F., & Reiss, K. (2014). Effects of collaboration scripts and heuristic worked examples on the acquisition of mathematical argumentation skills of teacher students with different levels of prior achievement. *Learning and Instruction, 32*, 22-36. doi:10.1016/j.learninstruc.2014.01.003

Koslowski, B. (1996). *Theory and Evidence: The Development of Scientific Reasoning*. LaVergne, TN: Lightning Source.

Kramarski, B., Mevarech, Z. R., & Arami, M. (2002). The effects of metacognitive instruction on solving mathematical authentic tasks. *Educational Studies in Mathematics, 49*(2), 225-250. doi:10.1023/A:1016282811724

Krathwohl, D. R. (2002). A revision of Bloom's taxonomy: An overview. *Theory into practice, 41*(4), 212-218.

Kühberger, A., Fritz, A., & Scherndl, T. (2014). Publication Bias in Psychology: A Diagnosis Based on the Correlation between Effect Size and Sample Size. *PLOS ONE, 9*(9), e105825. doi:10.1371/journal.pone.0105825

Kuhn, D. (2002). What is scientific thinking, and how does it develop?. In U. Goswami (Ed.), *Blackwell handbook of childhood cognitive development* (pp. 371-393). New York, NY: Blackwell Publishing.

Kuhn, D. (2009). Do students need to be taught how to reason? *Educational Research Review, 4*(1), 1-6. doi:10.1016/j.edurev.2008.11.001

Kuhn, D., Schauble, L., & Garcia-Mila, M. (1992). Cross-domain development of scientific reasoning. *Cognition and Instruction, 9*(4), 285-327.

Kuntze, S. (2008). Fostering geometrical proof competency by student-centred writing activities. In O. Figueras, J. L. Corina, S. Alatorre, T. Rojano, & A. Sepúlveda (Eds.), *Proceedings of the 32nd conference of the international group for the psychology of mathematics education* (Vol. 3, pp. 289-296). Morelia, México PME.

Lai, Y., & Weber, K. (2014). Factors mathematicians profess to consider when presenting pedagogical proofs. *Educational Studies in Mathematics, 85*(1), 93-108. doi:10.1007/s10649-013-9497-z

Lai, Y., Weber, K., & Mejía-Ramos, J. P. (2012). Mathematicians' Perspectives on Features of a Good Pedagogical Proof. *Cognition and Instruction, 30*(2), 146-169. doi:10.1080/07370008.2012.661814

Lakatos, I. (1976). *Proofs and Refutations*. Cambridge, UK: Cambridge University Press.

Lakoff, G. (1990). *Women, fire, and dangerous things: What categories reveal about the mind*. Cambridge, UK: Cambridge Univ Press.

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge, UK: Cambridge University Press.

Leder, G. C., Pehkonen, E., & Törner, G. (2002). *Beliefs: A hidden variable in mathematics education?* (Vol. 31). Dordrecht, Netherlands: Kluwer Academic Publishers.

Lehmann, D., & Magidor, M. (1992). What does a conditional knowledge base entail? *Artificial intelligence, 55*(1), 1-60.

Leighton, J. P. (2006). Teaching and Assessing Deductive Reasoning Skills. *The Journal of Experimental Education, 74*(2), 109-136.

Leighton, J. P., & Sternberg, R. J. (2003). Reasoning and Problem Solving. In A. F. Healy, R. W. Proctor, & I. B. Weiner (Eds.), *Handbook of Psychology*. Hoboken, NJ: John Wiley & Sons.

Leighton, J. P., & Sternberg, R. J. (2004). *The nature of reasoning*. Cambridge, UK: Cambridge University Press

Leiss, D., & Blum, W. (2006). Beschreibung zentraler mathematischer Kompetenzen. In W. Blum, C. Drüke-Noe, R. Hartung, & O. Köller (Eds.), *Bildungsstandards Mathematik: konkret. Sekundarstufe I: Aufgabenbeispiele, Unterrichtsanregungen, Fortbildungsideen*. Berlin, Germany Cornelsen.

Levie, W. H., & Lentz, R. (1982). Effects of text illustrations: A review of research. *Educational Communication and Technology Journal, 30*(4), 195-232.

Lew, K., Fukawa-Connelly, T., Mejía-Ramos, J. P., & Weber, K. (2016). Lectures in Advanced Mathematics: Why Students Might Not Understand What the Mathematics Professor Is Trying to Convey. *Journal for Research in Mathematics Education, 47*(2), 162-198.

Lim, J. H. (2006). *Effects of part-task and whole-task instructional approaches and learner levels of expertise on learner performance of a complex cognitive task.* Florida State University, Tallahassee. FL.

Lim, J. H., Reiser, R. A., & Olina, Z. (2009). The effects of part-task and whole-task instructional approaches on acquisition and transfer of a complex cognitive skill. *Educational Technology Research and Development, 57*(1), 61-77. doi:10.1007/s11423-007-9085-y

Lin, F.-L. (2005). Modeling students' learning on mathematical proof and refutation. In H. L. Chick & J. L. Vincent (Eds.), *Proceedings of the 29th Conference of the International Group for the Psychology of Mathematics Education* (Vol. 1, pp. 3-18). Melbourne, Australia.

Lin, F.-L., & Yang, K.-L. (2007). The Reading Comprehension of Geometric Proofs: The Contribution of Knowledge and Reasoning. *International Journal of Science and Mathematics Education, 5*(4), 729-754. doi:10.1007/s10763-007-9095-6

Lin, F.-L., Yang, K.-L., Lee, K.-H., Tabach, M., & Stylianides, G. (2012). Principles of Task Design for Conjecturing and Proving. In G. Hanna & M. de Villiers (Eds.), *Proof and Proving in Mathematics Education: The 19th ICMI Study* (pp. 305-325). Dordrecht, Netherlands: Springer Netherlands.

Lingel, K., Neuenhaus, N., Artelt, C., & Schneider, W. (2014). Der Einfluss des metakognitiven Wissens auf die Entwicklung der Mathematikleistung am Beginn der Sekundarstufe I. *Journal für Mathematik-Didaktik, 35*(1), 1-29. doi:10.1007/s13138-013-0061-2

Lithner, J. (2008). A research framework for creative and imitative reasoning. *Educational Studies in Mathematics, 67*(3), 255-276. doi:10.1007/s10649-007-9104-2

Livermore, A. H. (1964). The process approach of the AAAS commission on science education. *Journal of Research in Science Teaching, 2*(4), 271-282. doi:10.1002/tea.3660020403

Lockwood, E., Ellis, A. B., & Lynch, A. G. (2016). Mathematicians' Example-Related Activity when Exploring and Proving Conjectures. *International Journal of Research in Undergraduate Mathematics Education, 2*(2), 165-196. doi:10.1007/s40753-016-0025-2

Looi, C.-K., & Teh, L. W. (2015). *Scaling Educational Innovations*. Singapore, Singapore: Springer Singapore.

Maciejewski, W., & Star, J. R. (2016). Developing flexible procedural knowledge in undergraduate calculus. *Research in Mathematics Education, 18*(3), 299-316. doi:10.1080/14794802.2016.1148626

Malek, A., & Movshovitz-Hadar, N. (2011). The effect of using Transparent Pseudo-Proofs in Linear Algebra. *Research in Mathematics Education, 13*(1), 33-58. doi:10.1080/14794802.2011.550719

Mamona-Downs, J., & Downs, M. (2005). The identity of problem solving. *The Journal of Mathematical Behavior, 24*(3-4), 385-401. doi:10.1016/j.jmathb.2005.09.011

Manin, Y. (2010). *A Course in Mathematical Logic for Mathematicians*. New York, NY: Springer US.

Mariotti, M. A. (2006). Proof and proving in mathematics education. In Á. Gutiérrez & P. Boero (Eds.), *Handbook of research on the psychology of mathematics education: Past, present and future* (pp. 173-204). Rotterdam, Netherlands: Sense Publishers.

Mason, J. (1982). *Thinking mathematically*. Boston, MA: Addison Wesley.

Matos, J. F. (2013). Trends and opportunities for research methodologies within the PME community. In A. Lindmeier & A. Heinze (Eds.), *Proceedings of the 37th Conference of the International Group for the Psychology of Mathematics Education* (Vol. 1, pp. 51-66). Kiel, Germany: PME.

Mayer, R. E. (1983). *Thinking, problem solving, cognition .* New York, NY: W.H. Freeman.

Mayer, R. E. (1998). Cognitive, metacognitive, and motivational aspects of problem solving. *Instructional Science, 26*(1-2), 49-63. doi:10.1023/A:1003088013286

Mayer, R. E. (2010). Applying the science of learning to medical education. *Medical Education, 44*(6), 543-549. doi:10.1111/j.1365-2923.2010.03624.x

Mayer, R. E., & Wittrock, M. C. (2006). Problem Solving. In P. Alexander & P. Winne (Eds.), *Handbook of educational psychology* (pp. 287-304). Mahwah, NJ: Lawrence Erlbaum Associates.

Mejía-Ramos, J. P., Fuller, E., Weber, K., Rhoads, K., & Samkoff, A. (2012). An assessment model for proof comprehension in undergraduate mathematics. *Educational Studies in Mathematics, 79*(1), 3-18. doi:10.1007/s10649-011-9349-7

Mejía-Ramos, J. P., & Inglis, M. (2009a). Argumentative and proving activities in mathematics education research. In F.-L. Lin, F.-J. Hsieh, G. Hanna, & M. de Villiers (Eds.), *Proceedings of the ICMI Study 19 conference: Proof and Proving in Mathematics Education* (Vol. 2, pp. 88-93). Taipei, Taiwan.

Mejía-Ramos, J. P., & Inglis, M. (2009b). What are the argumentative activities associated with proof? *Research in Mathematics Education, 11*(1), 77-78. doi:10.1080/14794800902732258

Mejía-Ramos, J. P., & Weber, K. (2014). Why and how mathematicians read proofs: further evidence from a survey study. *Educational Studies in Mathematics, 85*(2), 161-173. doi:10.1007/s10649-013-9514-2

Mejía-Ramos, J. P., Weber, K., & Fuller, E. (2015). Factors Influencing Students' Propensity for Semantic and Syntactic Reasoning in Proof Writing: a Case Study. *International Journal of Research in Undergraduate Mathematics Education, 1*(2), 187-208. doi:10.1007/s40753-015-0014-x

Merrill, M. D. (2002). A pebble-in-the-pond model for instructional design. *Performance improvement, 41*(7), 41-46.

Mevarech, Z., & Fridkin, S. (2006). The effects of IMPROVE on mathematical knowledge, mathematical reasoning and meta-cognition. *Metacognition and Learning, 1*(1), 85-97. doi:10.1007/s11409-006-6584-x

Meyer, M. (2007). *Entdecken und Begründen im Mathematikunterricht: von der Abduktion zum Argument*. Hildesheim, Germany: Franzbecker.

Möckel, T., Beste, C., & Wascher, E. (2015). The Effects of Time on Task in Response Selection - An ERP Study of Mental Fatigue. *Scientific Reports, 5*, 10113. doi:10.1038/srep10113

Mokhtari, K., & Reichard, C. A. (2002). Assessing students' metacognitive awareness of reading strategies. *Journal of educational psychology, 94*(2), 249-259.

Moore, R. C. (1994). Making the transition to formal proof. *Educational Studies in Mathematics, 27*(3), 249-266. doi:10.1007/bf01273731

Morris, A. K. (2007). Factors Affecting Pre-Service Teachers' Evaluations of the Validity of Students' Mathematical Arguments in Classroom Contexts. *Cognition and Instruction, 25*(4), 479-522. doi:10.1080/07370000701632405

Mulder, M., Gulikers, J., Biemans, H., & Wesselink, R. (2009). The new competence concept in higher education: error or enrichment? *Journal of European Industrial Training, 33*(8/9), 755-770. doi:10.1108/03090590910993616

Müller-Hill, E., & Kempen, L. (in preparation). Some suggestions on the enculturation function of mathematical proof.

National Council of Teachers of Mathematics. (2000). *Principles and Standards for School Mathematics*. Reston, VA: National Council of Teachers of Mathematics.

National Research Council. (2012). *Education for life and work: Developing transferable knowledge and skills in the 21st century*. Washington, DC: The National Academy Press.

Naylor, J. C., & Briggs, G. E. (1963). Effects of task complexity and task organization on the relative efficiency of part and whole training methods. *Journal of experimental psychology, 65*(3), 217-224.

Nelsen, R. B. (1993). *Proofs without words: Exercises in visual thinking*. Washington, DC: Mathematical Association of America.

Newell, A. (1980). Reasoning, problem solving and decision processes: The problem space as a fundamental category. In R. S. Nickerson (Ed.), *Attention and Performance VIII* (pp. 693-718). Hillsdale, NJ: Lawrence Erlbaum Associates.

Newell, A., & Simon, H. A. (1972). *Human problem solving* (Vol. 104). Englewood Cliffs, NJ: Prentice-Hall

Nickerson, S., & Rasmussen, C. (2009). Enculturation to proof: A pragmatic and theoretical investigation. In F.-L. Lin, F.-J. Hsieh, G. Hanna, & M. d. Villiers (Eds.), *Proceedings of the ICMI study 19 conference: Proof and proving in Mathematics Education* (pp. 118-123). Taipei, Taiwan.

Nunokawa, K. (1994). Solver's structures of a problem situation and their global restructing. *The Journal of Mathematical Behavior, 13*(3), 275 - 297. doi:10.1016/0732-3123(94)90009-4

Nussbaum, E. M. (2008). Collaborative discourse, argumentation, and learning: Preface and literature review. *Contemporary Educational Psychology, 33*, 345-359.

Obama, B. (2009). President Obama launches "Educate to Innovate" campaign for excellence in science, technology, engineering & math (STEM) education. Retrieved from the White House website: http://www.whitehouse.gov/the-press-office/presidentobama-launches-educate-innovate-campaign-excellence-science-technology-en.

OECD. (2013). *OECD Skills Outlook 2013: First Results from the Survey of Adult Skills*. Paris, France: OECD Publishing.

OECD. (2014). *PISA 2012 results: what students know and can do: Student performance in mathematics, reading and science (Volume 1, revised edition, February 2014)*. Paris, France: OECD Publishing.

Op ’t Eynde, P., Corte, E., & Verschaffel, L. (2002). Framing Students' Mathematics-Related Beliefs. In G. C. Leder, E. Pehkonen, & G. Törner (Eds.), *Beliefs: A Hidden Variable in Mathematics Education?* (pp. 13-37). Dordrecht, Netherlands: Springer Netherlands.

Opitz, A., Heene, M., & Fischer, F. (2015). *If scientific reasoning is what the tests test, then what is scientific reasoning? A review.* Paper presented at the 2015 Annual Meeting - American Educational Research Association, Chicago, IL.

Osborne, J. (2013). The 21st century challenge for science education: Assessing scientific reasoning. *Thinking Skills and Creativity, 10*, 265 - 279. doi:10.1016/j.tsc.2013.07.006

Osborne, J., & Patterson, A. (2011). Scientific argument and explanation: A necessary distinction? *Science Education, 95*(4), 627-638. doi:10.1002/sce.20438

Ottinger, S., Kollar, I., & Ufer, S. (2016). Content and Form - All the Same or Different Qualities of Mathematical Arguments? In C. Csíkos, A. Rausch, & J. Szitányi (Eds.), *Proceedings of the 40th Conference of the International Group for the Psychology of Mathematics Education* (Vol. 4, pp. 19-26): Szeged, Hungary: PME.

Ottinger, S., Kollar, I., & Ufer, S. (2017). Mathematisches Argumentieren und Beweisen - Struktur individuell-kognitiver und sozial-diskursiver Prozessmerkmale. In Institut für Mathematik der Universität Potsdam (Ed.), *Beiträge zum Mathematikunterricht 2017*. Münster, Germany: WTM-Verlag.

Paris, S. G., Lipson, M. Y., & Wixson, K. K. (1983). Becoming a strategic reader. *Contemporary Educational Psychology, 8*(3), 293-316. doi:10.1016/0361-476X(83)90018-8

Pedemonte, B. (2007). How can the relationship between argumentation and proof be analysed? *Educational Studies in Mathematics, 66*(1), 23-41. doi:10.1007/s10649-006-9057-x

Pedemonte, B. (2008). Argumentation and algebraic proof. *Zentralblatt für Didaktik der Mathematik, 40*(3), 385-400. doi:10.1007/s11858-008-0085-0

Pekrun, R. (1992). The Impact of Emotions on Learning and Achievement: Towards a Theory of Cognitive/Motivational Mediators. *Applied Psychology, 41*(4), 359-376. doi:10.1111/j.1464-0597.1992.tb00712.x

Pekrun, R. (2006). The Control-Value Theory of Achievement Emotions: Assumptions, Corollaries, and Implications for Educational Research and Practice. *Educational Psychology Review, 18*(4), 315-341. doi:10.1007/s10648-006-9029-9

Pekrun, R., Goetz, T., Titz, W., & Perry, R. P. (2002). Academic Emotions in Students' Self-Regulated Learning and Achievement: A Program of Qualitative and Quantitative Research. *Educational Psychologist, 37*(2), 91-105. doi:10.1207/S15326985EP3702_4

Pekrun, R., & Stephens, E. J. (2012). Academic emotions. In K. R. Harris, S. Graham, & T. Urdan (Eds.), *APA educational psychology handbook* (Vol. 2): Washington, DC: American Psychological Association.

Perels, F., Gürtler, T., & Schmitz, B. (2005). Training of self-regulatory and problem-solving competence. *Learning and Instruction, 15*(2), 123-139. doi:10.1016/j.learninstruc.2005.04.010

Perkins, D. N., & Grotzer, T. A. (1997). Teaching intelligence. *American Psychologist, 52*(10), 1125-1133.

Perry, P., Samper, C., Camargo, L., Molina, Ó., & Echeverry, A. (2009). Learning to prove: enculturation or…? In F.-L. En Lin, F.-J. Hsieh, & G. Hanna (Eds.), *Proceedings of the ICMI Study 19 Conference: Proof and proving in mathematics education* (pp. 124-129). Taipei, Taiwan.

Pfeiffer, K. (2009a). The role of proof validation in students' mathematical learning. In S. Close, D. Corcoran, R. Ward, & T. Dooley (Eds.), *Proceedings of the third national conference on* *research in mathematics education (MEI 3)* (pp. 403-413). Dublin, Ireland: St. Patricks College.

Pfeiffer, K. (2009b). The role of proof validation in students' mathematical learning. In M. Joubert (Ed.), *Proceedings of the British Society for Research into Learning Mathematics* (Vol. 29(3), pp. 79-84). Bristol, UK: British Society for Research into Learning Mathematics.

Pfeiffer, K. (2010). How do students evaluate and compare mathematical proofs? *Research in Mathematics Education, 12*(2), 161-162. doi:10.1080/14794802.2010.496986

Pfeiffer, K. (2011). *Features and purposes of mathematical proofs in the view of novice students: Observations from proof validation and evaluation performances.* National University of Ireland, Galway.

Pfeiffer, K., & Quinlan, R. (2015). Proof evaluation tasks as tools for teaching? In K. Krainer & N. a. Vondrová (Eds.), *Proceedings of the CERME 9 - Ninth Congress of the European Society for Research in Mathematics Education* (pp. 178–184). Prague, Czech Republic.

Philipp, K. (2012). *Experimentelles Denken: theoretische und empirische Konkretisierung einer mathematischen Kompetenz*. Wiesbaden, Germany: Springer Spektrum.

Pinto, M., & Tall, D. (1999). Student constructions of formal theory: giving and extracting meaning. In O. Zaslavsky (Ed.), *Proceedings of the 23rd Conference of the International Group for the Psychology of Mathematics Education* (Vol. 4, pp. 4-65). Haifa, Israel: PME.

Polya, G. (1945). *How to Solve it: A New Aspect of Mathematica Method*. Princeton, NJ: Princeton University Press.

Powers, R. A., Craviotto, C., & Grassl, R. M. (2010). Impact of proof validation on proof writing in abstract algebra. *International Journal of Mathematical Education in Science and Technology, 41*(4), 501-514. doi:10.1080/00207390903564603

Rach, S., & Heinze, A. (2011). Studying mathematics at the university: the influence of learning strategies. In B. Ubuz (Ed.), *Proceedings of the 35th conference of the international group for the psychology of mathematics education* (Vol. 4, pp. 9-16). Ankara, Turkey.

Rach, S., & Heinze, A. (2013). Welche Studierenden sind im ersten Semester erfolgreich? *Journal für Mathematik-Didaktik, 34*(1), 121-147. doi:10.1007/s13138-012-0049-3

Rach, S., & Heinze, A. (2016). The Transition from School to University in Mathematics: Which Influence Do School-Related Variables Have? *International Journal of Science and Mathematics Education*. doi:10.1007/s10763-016-9744-8

Rach, S., Heinze, A., & Ufer, S. (2014). Welche mathematischen Anforderungen erwarten Studierende im ersten Semester des Mathematikstudiums? *Journal für Mathematik-Didaktik, 35*(2), 205-228. doi:10.1007/s13138-014-0064-7

Raman, M. (2003). Key ideas: what are they and how can they help us understand how people view proof? *Educational Studies in Mathematics, 52*(3), 319-325.

Reichersdorfer, E. (2013). *Unterstützungsmaßnahmen am Beginn des Mathematikstudiums: Heuristische Lösungsbeispiele und Problemlösen in problembasierten Lernumgebungen zur Förderung mathematischer Argumentationskompetenz.* München, Germany: Universitätsbibliothek der TU München.

Reichersdorfer, E., Ufer, S., Lindmeier, A., & Reiss, K. (2014). The Transition from School to University: Theoretical Foundation and Practical Implementation of a Support Action at the Beginning of the Mathematics Studies. In I. Bausch, R. Biehler, R. Bruder, P. R. Fischer, R. Hochmuth, W. Koepf, S. Schreiber, & T. Wassong (Eds.), *Mathematische Vor- und Brückenkurse: Konzepte, Probleme und Perspektiven* (pp. 37-53). Wiesbaden, Germany: Springer Spektrum.

Reichersdorfer, E., Vogel, F., Fischer, F., Kollar, I., Reiss, K., & Ufer, S. (2012). Different collaborative learning settings to foster mathematical argumentation skills. In T.-Y. Tso

(Ed.), *Proceedings of the 36th Conference of the International Group for the Psychology of Mathematics Education* (Vol. 3, pp. 345-352). Taipei, Taiwan: PME.

Reid, D. A., & Knipping, C. (2010). *Proof in Mathematics Education. Research, Learning and Teaching*. Rotterdam, Netherlands: Sense Publishers.

Reiss, K., & Heinze, A. (2004). Knowledge acquisition in students' argumentation and proof processes. In G. Törner, N. Neill, A. Peter-Koop, & B. Wollring (Eds.), *Developments in Mathematic Education in Germany. Selected Papers from the Annual Conference on Didactics of Mathematics, Ludwigsburg* (pp. 107-115). Hildesheim, Deutschland Franzbecker.

Reiss, K., Heinze, A., & Kessler, S. (2007). Fostering argumentation and proof competencies in the mathematics classroom. In M. Prenzel (Ed.), *Studies on the educational quality of schools* (pp. 251-264). Münster, Deutschland: Waxmann.

Reiss, K., Heinze, A., & Klieme, E. (2002). Argumentation, proof and the understanding of proof. In H.-G. Weigand, N. Neill, A. Peter-Koop, K. Reiss, G. Törner, & B. Wollring (Eds.), *Developments in Mathematics Education in German-speaking Countries. Selected Papers from the Annual Conference on Didactics of Mathematics, Ludwigsburg* (pp. 109-120). Potsdam, Germany: Franzbecker

Reiss, K., Heinze, A., Renkl, A., & Groß, C. (2008). Reasoning and proof in geometry: Effects of a learning environment based on heuristic worked-out examples. *Zentralblatt für Didaktik der Mathematik, 40*(3), 455-467.

Reiss, K., & Renkl, A. (2002). Learning to prove: The idea of heuristic examples. *Zentralblatt für Didaktik der Mathematik, 34*(1), 29-35.

Reiss, K., & Ufer, S. (2009). Was macht mathematisches Arbeiten aus? Empirische Ergebnisse zum Argumentieren, Begründen und Beweisen. *Jahresbericht der Deutschen Mathematiker-Vereinigung (DMV), 111*(4), 155-177.

Renkl, A., Hilbert, T., & Schworm, S. (2009). Example-Based Learning in Heuristic Domains: A Cognitive Load Theory Account. *Educational Psychology Review, 21*(1), 67-78. doi:10.1007/s10648-008-9093-4

Richey, J. E., & Nokes-Malach, T. J. (2014). Comparing Four Instructional Techniques for Promoting Robust Knowledge. *Educational Psychology Review, 27*(1), 181-218.

Rittle-Johnson, B., & Schneider, M. (2014). Developing Conceptual and Procedural Knowledge of Mathematics. In R. Cohen Kadosh & A. Dowker (Eds.), *Oxford handbook of numerical cognition* (Vol. 10). Oxford, UK: Oxford university press.

Rittle-Johnson, B., Schneider, M., & Star, J. R. (2015). Not a One-Way Street: Bidirectional Relations Between Procedural and Conceptual Knowledge of Mathematics. *Educational Psychology Review, 27*(4), 587-597.

Rittle-Johnson, B., Siegler, R. S., & Alibali, M. W. (2001). Developing conceptual understanding and procedural skill in mathematics: An iterative process. *Journal of educational psychology, 93*(2), 346-362.

Roesken, B., Pepin, B., & Toerner, G. (2011). Beliefs and beyond: affect and the teaching and learning of mathematics. *Zentralblatt für Didaktik der Mathematik, 43*(4), 451-455. doi:10.1007/s11858-011-0354-1

Rowland, T. (2001). Generic proofs in number theory. In S. Campbell & R. Zazkis (Eds.), *Learning and teaching number theory: Research in cognition and instruction* (pp. 157-183). Westport, CT: Ablex.

Rowland, T. (2013). The Knowledge Quartet: the genesis and application of a framework for analysing mathematics teaching and deepening teachers' mathematics knowledge. *Sisyphus-Journal of Education, 1*(3), 15-43.

Roy, S., Alcock, L., & Inglis, M. (2010). Undergraduates proof comprehension: A comparative study of three forms of proof presentation *Proceedings of the 13th Conference for Research in Undergraduate Mathematics Education*. Raleigh, NC.
Sadler, D. R. (2013). Making Competent Judgments of Competence. In S. Blömeke, O. Zlatkin-Troitschanskaia, C. Kuhn, & J. Fege (Eds.), *Modeling and Measuring Competencies in Higher Education: Tasks and Challenges* (pp. 13-27). Rotterdam, Netherlands: SensePublishers.
Salden, R. J. C. M., Aleven, V., Schwonke, R., & Renkl, A. (2010). The expertise reversal effect and worked examples in tutored problem solving. *Instructional Science, 38*(3), 289-307. doi:10.1007/s11251-009-9107-8
Salden, R. J. C. M., Paas, F., & van Merriënboer, J. J. G. (2006). A comparison of approaches to learning task selection in the training of complex cognitive skills. *Computers in Human Behavior, 22*(3), 321-333. doi:10.1016/j.chb.2004.06.003
Samkoff, A., & Weber, K. (2015). Lessons learned from an instructional intervention on proof comprehension. *The Journal of Mathematical Behavior, 39*, 28 - 50. doi:10.1016/j.jmathb.2015.05.002
Samper, C., Camargo, L., Perry, P., & Molina, Î. (2012). Dynamic geometry, implication and abduction: A case study. In T.-Y. Tso (Ed.), *Proceedings of the 36th Conference of the International Group for the Psychology of Mathematics Education* (Vol. 4, pp. 43-49). Taipeh, Taiwan: PME.
Sandefur, J., Mason, J., Stylianides, G. J., & Watson, A. (2013). Generating and using examples in the proving process. *Educational Studies in Mathematics, 83*(3), 323-340. doi:10.1007/s10649-012-9459-x
Scherer, R. (2015). Is it time for a new measurement approach? A closer look at the assessment of cognitive adaptability in complex problem solving. *Frontiers in psychology, 6*, 1664-1667.
Schneider, M. (2006). *Konzeptuelles und prozedurales Wissen als latente Variablen: Ihre Interaktion beim Lernen mit Dezimalbrüchen.* Technische Universität Berlin.
Schoenfeld, A. H. (1982). Measures of problem-solving performance and of problem-solving instruction. *Journal for Research in Mathematics Education, 13*(1), 31-49.
Schoenfeld, A. H. (1983). Beyond the Purely Cognitive: Belief Systems, Social Cognitions, and Metacognitions As Driving Forces in Intellectual Performance. *Cognitive Science, 7*(4), 329-363. doi:10.1207/s15516709cog0704_3
Schoenfeld, A. H. (1985). *Mathematical problem solving*. New York, NY: Academic press.
Schoenfeld, A. H. (1987). What's All the Fuss About Metacognitlon. In A. H. Schoenfeld (Ed.), *Cognitive science and mathematics education* (pp. 189-215). Hillsdale, NJ: Lawrence Erlbaum Asociates
Schoenfeld, A. H. (1989). Teaching mathematical thinking and problem solving. In B. R. Joyce, L. B. Resnick, & L. E. Klopfer (Eds.), *Toward the thinking curriculum: Current cognitive research* (pp. 83-103). Alexandria, VI: Association for Supervision and Curriculum Development.
Schoenfeld, A. H. (1992). Learning to Think Mathematically: Problem Solving, metacognition, and Sense-Making in Mathematics In D. Grouws (Ed.), *The Handbook for Research on Mathematics Teaching and Learning* (pp. 334-370). New York, NY: MacMillan.
Schoenfeld, A. H. (2010). *How we think: a theory of goal-oriented decision making and its educational applications*. New York, NY: Routledge.

Schoenfeld, A. H. (2012a). How We Think: A Theory of Human Decision-making, with a Focus on Teaching. In S. J. Cho (Ed.), *Proceedings of the 12th International Congress on Mathematical Education* (pp. 229-243). Seoul, Korea.

Schoenfeld, A. H. (2012b). Problematizing the didactic triangle. *Zentralblatt für Didaktik der Mathematik, 44*(5), 587-599. doi:10.1007/s11858-012-0395-0

Schraw, G. (1998). Promoting general metacognitive awareness. *Instructional Science, 26*(1), 113-125. doi:10.1023/A:1003044231033

Schraw, G., Crippen, K., & Hartley, K. (2006). Promoting Self-Regulation in Science Education: Metacognition as Part of a Broader Perspective on Learning. *Research in Science Education, 36*(1-2), 111-139. doi:10.1007/s11165-005-3917-8

Schraw, G., & Dennison, R. S. (1994). Assessing metacognitive awareness. *Contemporary educational psychology, 19*(4), 460-475.

Schunn, C. D., & Anderson, J. R. (1999). The Generality/Specificity of Expertise in Scientific Reasoning. *Cognitive Science, 23*(3), 337-370. doi:10.1207/s15516709cog2303_3

Schwarz, B. B., Hershkowitz, R., & Prusak, N. (2010). Argumentation and mathematics. In K. Littleton & C. Howe (Eds.), *Educational dialogues: Understanding and promoting productive interaction* (pp. 115-141). New York, NY: Routledge.

Selden, A. (2011). Transitions and proof and proving at tertiary level. In G. Hanna & M. de Villiers (Eds.), *Proof and proving in mathematics education* (pp. 391-420). Dordrecht, Netherlands: Springer Netherlands.

Selden, A., & Selden, J. (1978). Errors students make in mathematical reasoning. *Bōgazici Üniversitesi Dergisi, 6*, 67-87.

Selden, A., & Selden, J. (2003). Validations of Proofs Considered as Texts: Can Undergraduates Tell Whether an Argument Proves a Theorem? *Journal for Research in Mathematics Education, 34*(1), 4-36.

Selden, A., & Selden, J. (2008). Overcoming students' difficulties in learning to understand and construct proofs. In M. P. Carlsson & C. Rasmussen (Eds.), *Making the Connection: Research and Practice in Undergraduate Mathematics* (pp. 95-110). Washington, DC: Mathematical Association of America.

Selden, A., & Selden, J. (2012). *Mathematical and non-mathematical university students' proving difficulties.* Paper presented at the 15th Annual Legacy of R. L. Moore Conference, Austin, TX.

Selden, A., & Selden, J. (2013). Proof and problem solving at university level. *The Mathematics Enthusiast, 10*(1-2), 303-334.

Selden, A., & Selden, J. (2015a). A comparison of proof comprehension, proof construction, proof validation and proof evaluation. In R. Göller, R. Biehler, R. Hochmuth, & H.-G. Rück (Eds.), *KHDM Proceedings Hannover Germany December 2015*. Hannover, Germany.

Selden, A., & Selden, J. (2015b). A theoretical perspective for proof construction. In K. Krainer & N. Vondrová (Eds.), *Proceedings of the CERME 9 - Ninth Congress of the European Society for Research in Mathematics Education* (pp. 198-204). Prague, Czech Republic.

Selden, J., & Selden, A. (1995). Unpacking the logic of mathematical statements. *Educational Studies in Mathematics, 29*(2), 123-151. doi:10.1007/BF01274210

Selden, J., & Selden, A. (2009). Understanding the proof construction process. In F.-L. Lin, F.-J. Hsieh, G. Hanna, & M. d. Villiers (Eds.), *Proceedings of the ICMI Study 19 Conference: Proof and proving in mathematics education* (Vol. 2, pp. 196-201). Taipei, Taiwan.

Senk, S. L. (1985). How Well Do Students Write Geometry Proofs? *The Mathematics Teacher, 78*(6), 448-456.

Senk, S. L. (1989). Van Hiele Levels and Achievement in Writing Geometry Proofs. *Journal for Research in Mathematics Education, 20*(3), 309-321. doi:10.2307/749519

Seymour, E., & Ferrare, J. J. (2015). Talking about Leaving Revisited: What do we know about why undergraduates leave the sciences? Retrieved from http://talr.wceruw.org/documents/GardnerInstitute2015FINAL.pdf

Seymour, E., & Hewitt, N. M. (1997). *Talking About Leaving: Why Undergraduates Leave the Sciences.* Boulder, CO: Westview Press.

Shavelson, R. J. (2010). On the measurement of competency. *Empirical research in vocational education and training, 2*(1), 41-63.

Sherin, M. G., & van Es, E. A. (2009). Effects of video club participation on teachers' professional vision. *Journal of Teacher Education, 60*(1), 20-37.

Shulman, L. S. (1986). Those Who Understand: Knowledge Growth in Teaching. *Educational Researcher, 15*(2), 4-14. doi:10.3102/0013189X015002004

Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard educational review, 57*(1), 1-23.

Siebert, U., Rach, S., & Heinze, A. (2013). Teaching quality of mathematics university courses. In A. Lindmeier & A. Heinze (Eds.), *Proceedings of the 37th Conference of the International Group for the Psychology of Mathematics Education* (Vol. 5, pp. 269). Kiel, Germany: PME.

Siegler, R. S. (1989). How Domain-General and Domain-Specific Knowledge Interact to Produce Strategy Choices. *Merrill-Palmer Quarterly, 35*(1), 1-26.

Simon, H. A. (1978). Information-processing theory of human problem solving. In W. K. Estes (Ed.), *Handbook of learning and cognitive processes* (Vol. 5, pp. 271-295). Hillsdale, NJ: Lawrence Erlabaum Associates.

Smith, J. C. (2006). A sense-making approach to proof: Strategies of students in traditional and problem-based number theory courses. *The Journal of Mathematical Behavior, 25*(1), 73-90. doi:10.1016/j.jmathb.2005.11.005

Snook, I., O’ Neill, J., Clark, J., O’ Neill, A.-M., & Openshaw, R. (2009). Invisible learnings? A commentary on John Hattie's book Visible Learning: a synthesis of over 800 meta-analyses relating to achievement. *New Zealand journal of educational studies, 44*(1), 93-106.

So, J. C. Y., Proctor, R. W., Dunston, P. S., & Wang, X. (2013). Better Retention of Skill Operating a Simulated Hydraulic Excavator After Part-Task Than After Whole-Task Training. *Human Factors, 55*(2), 449-460. doi:10.1177/0018720812454292

Sodian, B., & Bullock, M. (2008). Scientific reasoning - Where are we now? *Cognitive Development, 4*, 431-434.

Sommerhoff, D., Ufer, S., & Kollar, I. (2015a). Forschung zum Mathematischen Argumentieren - Ein deskriptiver Review von PME-Beiträgen. In F. Caluori, H. Linneweber-Lammerskitten, & C. Streit (Eds.), *Beiträge zum Mathematikunterricht 2015*. Münster, Germany: WTM-Verlag.

Sommerhoff, D., Ufer, S., & Kollar, I. (2015b). Research on mathematical argumentation: A descriptive review of PME proceedings. In K. Beswick, T. Muir, & J. Wells (Eds.), *Proceedings of the 39th Conference of the International Group for the Psychology of Mathematics Education* (Vol. 4, pp. 193-200). Hobart, Australia: PME.

Sommerhoff, D., Ufer, S., & Kollar, I. (2016a). Mathematical Argumentation and Proof - Supporting a Complex Cognitive Skill. In C. K. Looi, J. L. Polman, U. Cress, & P. Reimann (Eds.), *Transforming Learning, Empowering Learners: The International Conference of the Learning Sciences* (Vol. 2, pp. 811-814). Singapore, Singapore: International Society of the Learning Sciences.

Sommerhoff, D., Ufer, S., & Kollar, I. (2016b). Proof Validation Aspects and Cognitive Student Prerequisites in Undergraduate Mathematics. In C. Csíkos, A. Rausch, & J. Szitányi (Eds.), *Proceedings of the 40th Conference of the International Group for the Psychology of Mathematics Education* (Vol. 4, pp. 219-226): Szeged, Hungary: PME.

Sommerhoff, D., Ufer, S., & Kollar, I. (2016c). Validieren von Beweisen - Probleme von Studierenden und die Rolle von mathematischen und übergreifenden Voraussetzungen. In I. für Mathematik und Informatik Heidelberg (Ed.), *Beiträge zum Mathematikunterricht 2016*. Münster, Germany: WTM-Verlag.

Stahnke, R., Schueler, S., & Roesken-Winter, B. (2016). Teachers' perception, interpretation, and decision-making: a systematic review of empirical mathematics education research. *Zentralblatt für Didaktik der Mathematik, 48*(1), 1-27. doi:10.1007/s11858-016-0775-y

Star, J. R. (2005). Reconceptualizing Procedural Knowledge. *Journal for Research in Mathematics Education, 36*(5), 404-411. doi:10.2307/30034943

Star, J. R., & Stylianides, G. J. (2013). Procedural and conceptual knowledge: exploring the gap between knowledge type and knowledge quality. *Canadian Journal of Science, Mathematics and Technology Education, 13*(2), 169-181.

Sternberg, R. J., Birney, D., Jarvin, L., Kirlik, A., Stemler, S., & Grigorenko, E. L. (2006). From molehill to mountain: The proecss of scaling up educational interventions (First hand experience upscaling the theory of successful intelligence). In R. J. Sternberg & M. A. Constas (Eds.), *Translating theory and research into educational practice: Developments in content domains, large-scale reform, and intellectual capacity* (pp. 205-222). Mahwah, NJ: Lawrence Erlbaum Associates.

Stokes, D. E. (1997). *Pasteur's quadrant : basic science and technological innovation*. Washington, D.C.: Brookings Institution Press.

Strijbos, J.-W., Martens, R. L., Prins, F. J., & Jochems, W. M. G. (2006). Content analysis: What are they talking about? *Computers & Education, 46*(1), 29-48. doi:10.1016/j.compedu.2005.04.002

Strijbos, J.-W., & Stahl, G. (2007). Methodological issues in developing a multi-dimensional coding procedure for small-group chat communication. *Learning and Instruction, 17*(4), 394-404. doi:10.1016/j.learninstruc.2007.03.005

Stylianides, A. J. (2007). Proof and Proving in School Mathematics. *Journal for Research in Mathematics Education, 38*(3), 289-321.

Stylianides, A. J., & Stylianides, G. J. (2006). Content Knowledge for Mathematics Teaching: The Case of Reasoning and Proving. In J. Novotná, H. Moraová, M. Krátká, & N. Stehlínková (Eds.), *Proceedings of the 30th Conference of the International Group for the Psychology of Mathematics Education* (Vol. 5, pp. 201-208). Prague, Hungary: PME.

Stylianides, G. J. (2008). An Analytic Framework of Reasoning-and-Proving. *For the Learning of Mathematics, 28*(1), 9-16.

Stylianides, G. J., Stylianides, A. J., & Weber, K. (in press). Research on the teaching and learning of proof: Taking stock and moving forward. In J. Cai (Ed.), *First compendium for research in mathematics education*. Reston, VA: National Council of Teachers of Mathematics.

Stylianou, D. A., Blanton, M. L., & Rotou, O. (2015). Undergraduate Students' Understanding of Proof: Relationships Between Proof Conceptions, Beliefs, and Classroom Experiences with Learning Proof. *International Journal of Research in Undergraduate Mathematics Education, 1*(1), 91-134. doi:10.1007/s40753-015-0003-0

Südkamp, A., & Praetorius, A.-K. (2017). *Diagnostische Kompetenz von Lehrkräften*. Münster, Germany: Waxmann.

Sweller, J. (1990). On the limited evidence for the effectiveness of teaching general problem-solving strategies. *Journal for Research in Mathematics Education, 21*(5), 411-415.

Tall, D. (1991). The psychology of advanced mathematical thinking. In D. Tall (Ed.), *Advanced mathematical thinking* (pp. 3-21). Dordrecht, Netherlands: Springer Science + Business Media.

Tall, D. (1992). The transition to advanced mathematical thinking: functions, limits, infinity and proof. In D. Grouws (Ed.), *Handbook of Research on Mathematics Teaching and Learning* (pp. 495-511). New York, NY: MacMillan.

Tall, D. (2008). The transition to formal thinking in mathematics. *Mathematics Education Research Journal, 20*(2), 5-24.

Tanswell, F. (2012). *Proof and Prejudice: Why Formalising doesn't make you a Formalist.* Universiteit van Amsterdam.

The Cognition and Technology Group At Vanderbilt. (1990). Anchored Instruction and Its Relationship to Situated Cognition. *Educational Researcher, 19*(6), 2-10. doi:10.3102/0013189X019006002

The Design-Based Research Collective. (2003). Design-based research: An emerging paradigm for educational inquiry. *Educational Researcher, 32*(1), 5-8.

The International Commission on Mathematical Instruction (ICMI). (1966). Report on the Conference "Mathematics at the coming to university: real situation and desirable situation" *New Trends in Mathematics Teaching* (Vol. 1). Paris, France: UNESCO.

Thurston, W. P. (1994). On proof and progress in mathematics. In T. Tymoczko (Ed.), *New directions in the Philosophy of Mathematics: an Anthology* (pp. 337-355). Princeton, NY: Princeton University Press.

Todd, P. M., & Gigerenzer, G. (2001). Précis of Simple heuristics that make us smart. *Behavioral and Brain Sciences, 23*(5), 727-741.

Törner, G. (2002). Mathematical Beliefs - A Search for a Common Ground: Some Theoretical Considerations on Structuring Beliefs, some Research Questions, and some Phenomenological Observations. In G. C. Leder, E. Pehkonen, & G. Törner (Eds.), *Beliefs: A Hidden Variable in Mathematics Education?* (pp. 73-94). New York, NY: Kluwer Academic Publishers.

Toulmin, S. E. (2003). *The Uses of Argument* (2nd ed.). Cambridge, UK: Cambridge University Press.

Trenholm, S., Alcock, L., & Robinson, C. (2016). The instructor experience of fully online tertiary mathematics: a challenge and an opportunity. *Journal for Research in Mathematics Education, 47*(2), 147-161.

Tricot, A., & Sweller, J. (2014). Domain-Specific Knowledge and Why Teaching Generic Skills Does Not Work. *Educational Psychology Review, 26*(2), 265-283. doi:10.1007/s10648-013-9243-1

Trilling, B., & Fadel, C. (2009). *21st century skills: Learning for life in our times*. Hoboken, NJ: John Wiley & Sons.

Ubuz, B., Dincer, S., & Bulbul, A. (2012). Argumentation in undergraduate math courses: A study on proof generation. In T.-Y. Tso (Ed.), *Proceedings of the 36th Conference of the International Group for the Psychology of Mathematics Education*. Taipei, Taiwan: PME.

Ufer, S., Heinze, A., Kuntze, S., & Rudolph-Albert, F. (2009). Beweisen und Begründen im Mathematikunterricht: Die Rolle von Methodenwissen für das Beweisen in der Geometrie. *Journal für Mathematik-Didaktik, 30*(1), 30-54.

Ufer, S., Heinze, A., & Reiss, K. (2008). Individual predictors of geometrical proof competence. In O. Figueras, J. L. Cortina, S. Alatorre, T. Rojano, & A. Sepúlveda (Eds.), *Proceedings of the*

Joint Meeting of the 32nd Conference of the International Group for the Psychology of Mathematics Education and the XX North American Chapter (Vol. 1, pp. 361-368). Morelia, Mexico: PME.

Ufer, S., Heinze, A., & Reiss, K. (2009a). Mental models and the development of geometric proof competency. In M. Tzekaki, M. Kaldrimidou, & H. Sakonidis (Eds.), *Proceedings of the 33rd conference of the international group for the psychology of mathematics education* (Vol. 5, pp. 257-264). Thessaloniki, Greece: PME.

Ufer, S., Heinze, A., & Reiss, K. (2009b). What happens in students' minds when constructing a geometric proof? A cognitive model based on mental models. In F.-L. Lin, F.-J. Hsieh, G. Hanna, & M. de Villiers (Eds.), *Proceedings of the ICMI Study 19 conference: Proof and Proving in Mathematics Education* (Vol. 2, pp. 239-244). Taipei, Taiwan.

van der Linden, D., Frese, M., & Meijman, T. F. (2003). Mental fatigue and the control of cognitive processes: effects on perseveration and planning. *Acta Psychologica, 113*(1), 45-65. doi:10.1016/S0001-6918(02)00150-6

van Eemeren, F. H., & Grootendorst, R. (1999). Developments in argumentation theory. In P. Coirier & J. E. B. Andriessen (Eds.), *Foundations if argumentative text processing* (pp. 43-57). Amsterdam, Netherlands: Amsterdam University Press.

van Gog, T., Paas, F., & van Merriënboer, J. J. G. (2004). Process-Oriented Worked Examples: Improving Transfer Performance Through Enhanced Understanding. *Instructional Science, 32*(1), 83-98. doi:10.1023/B:TRUC.0000021810.70784.b0

van Hiele, P., & van Hiele-Geldorf, D. (1978). Die Bedeutung der Denkebenen im Unterrichtssystem nach der didaktischen Methode. In H. Steiner (Ed.), *Didaktik der Mathematik*. Darmstadt, Germany: Wissenschaftliche Buchgesellschaft.

van Joolingen, W. R., & de Jong, T. (1997). An extended dual search space model of scientific discovery learning. *Instructional Science, 25*(5), 307-346.

van Merriënboer, J. J. G. (1997). *Training Complex Cognitive Skills: A Four-Component Instructional Design Model for Technical Training*. Englewood Cliffs, NJ: Educational Technology Publications.

van Merriënboer, J. J. G. (2002). Complex Learning. In N. M. Seel (Ed.), *Encyclopedia of the Sciences of Learning* (pp. 681 - 682). Dordrecht, Netherlands: Springer Science+Business Media.

van Merriënboer, J. J. G. (2013). Perspectives on problem solving and instruction. *Computers & Education, 64*, 153-160. doi:10.1016/j.compedu.2012.11.025

van Merriënboer, J. J. G., de Croock, M. B. M., & Jelsma, O. (1997). The transfer paradox: Effects of contextual interference on retention and transfer performance of a complex cognitive skill. *Perceptual and Motor Skills, 84*(3), 784-786.

van Merriënboer, J. J. G., Jelsma, O., & Paas, F. G. W. C. (1992). Training for reflective expertise: A four-component instructional design model for complex cognitive skills. *Educational Technology Research and Development, 40*(2), 23-43. doi:10.1007/BF02297047

van Merriënboer, J. J. G., & Kester, L. (2007). Whole-task models in education. In J. M. Spector, M. D. Merrill, J. v. Merriënboer, & M. P. Driscoll (Eds.), *Handbook of research on educational communications and technology* (pp. 441-456). New York, NY: Routledge.

van Merriënboer, J. J. G., & Kirschner, P. A. (2007). *Ten Steps to Complex Learning: A Systematic Approach to Four-component Instructional Design*. Mahwah, NJ: Lawrence Erlbaum Associates.

van Merriënboer, J. J. G., Kirschner, P. A., & Kester, L. (2003). Taking the Load Off a Learner's Mind: Instructional Design for Complex Learning. *Educational Psychologist, 38*(1), 5-13.

Van Spronsen, H. D. (2008). *Proof Processes of Novice Mathematics Proof Writers.* University of Montana.

VanLehn, K. (1996). Cognitive skill acquisition. *Annual review of psychology, 47*(1), 513-539.

Vogel, F., Kollar, I., Ufer, S., Reichersdorfer, E., Reiss, K., & Fischer, F. (2016). Developing argumentation skills in mathematics through computer-supported collaborative learning: the role of transactivity. *Instructional Science, 44*(5), 477-500. doi:10.1007/s11251-016-9380-2

Voss, J. F., Tyler, S. W., Yengo, L. A., & others. (1983). Individual differences in the solving of social science problems. *Individual differences in cognition, 1*, 205-232.

Voss, J. F., & Van Dyke, J. A. (2001). Argumentation in Psychology: Background Comments. *Discourse Processes, 32*(2-3), 89-111. doi:10.1080/0163853X.2001.9651593

Voss, J. F., Wiley, J., & Carretero, M. (1995). Acquiring intellectual skills. *Annual Review of Psychology, 46*, 155-181.

Wagner, D. (2011). Mathematische Kompetenzanforderungen in Schule und Hochschule: Die Rolle des formal-abstrahierenden Denkens. In R. Haug & L. Holzäpfel (Eds.), *Beiträge zum Mathematikunterricht 2011* (pp. 879-882). Münster, Germany: WTM-Verlag.

Watson, A., Sandefur, J., Mason, J., & Stylianides, G. (2013). The use of examples to provide representations in proving. In A. M. Lindmeier & A. Heinze (Eds.), *Proceedings of the 37th Conference of the International Group for the Psychology of Mathematics Education*. Kiel, Germany: PME.

Weber, K. (2001). Student difficulty in constructing proofs: The need for strategic knowledge. *Educational Studies in Mathematics, 48*(1), 101-119. doi:10.1023/A:1015535614355

Weber, K. (2003). Students' difficulties with proof. In A. Selden & J. Selden (Eds.), *The Mathematical Association of America Online: Research Sampler* (Vol. 8).

Weber, K. (2004). A framework for describing the processes that undergraduates use to construct proofs. In M. J. Høines & A. B. Fuglestad (Eds.), *Proceedings of the 28th Conference of the International Group for the Psychology of Mathematics Education* (Vol. 4, pp. 425-432). Bergen, Norway: PME.

Weber, K. (2005). Problem-solving, proving, and learning: The relationship between problem-solving processes and learning opportunities in the activity of proof construction. *The Journal of Mathematical Behavior, 24*(3-4), 351-360. doi:10.1016/j.jmathb.2005.09.005

Weber, K. (2008). How Mathematicians Determine If an Argument Is a Valid Proof. *Journal for Research in Mathematics Education, 39*(4), 431-459.

Weber, K. (2009). How syntactic reasoners can develop understanding, evaluate conjectures, and generate counterexamples in advanced mathematics. *The Journal of Mathematical Behavior, 28*(2), 200-208.

Weber, K. (2012). Mathematicians' perspectives on their pedagogical practice with respect to proof. *International Journal of Mathematical Education in Science and Technology, 43*(4), 463-482. doi:10.1080/0020739X.2011.622803

Weber, K. (2014). Proof as a cluster concept. In C. Nicol, S. Oesterle, P. Liljedahl, & D. Allan (Eds.), *Proceedings of the 38th Conference of the International Group for the Psychology of Mathematics Education and the 36th Conference of the North American Chapter of the Psychology of Mathematics Education* (Vol. 5, pp. 353-360). Vancouver, Canada: PME.

Weber, K. (2015). Effective Proof Reading Strategies for Comprehending Mathematical Proofs. *International Journal of Research in Undergraduate Mathematics Education, 1*(3), 289-314. doi:10.1007/s40753-015-0011-0

Weber, K., & Alcock, L. (2004). Semantic and Syntactic Proof Productions. *Educational Studies in Mathematics, 56*(2), 209-234. doi:10.1023/B:EDUC.0000040410.57253.a1

Weber, K., & Alcock, L. (2005). Using Warranted Implications to Understand and Validate Proofs. *For the Learning of Mathematics, 25*(1), 34-51.

Weber, K., & Alcock, L. (2009). Proof in advanced mathematics classes: semantic and syntactic reasoning in the representation system of proof. In D. A. Stylianou, M. L. Blanton, & E. J. Knuth (Eds.), *Teaching and Learning Proof Across the Grades: a K-16 Perspective* (pp. 323-338). New York, NY: Routledge.

Weber, K., Inglis, M., & Mejía-Ramos, J. P. (2014). How Mathematicians Obtain Conviction: Implications for Mathematics Instruction and Research on Epistemic Cognition. *Educational Psychologist, 49*(1), 36-58. doi:10.1080/00461520.2013.865527

Weber, K., & Mejia-ramos, J. P. (2014). Mathematics majors' beliefs about proof reading. *International Journal of Mathematical Education in Science and Technology, 45*(1), 89-103. doi:10.1080/0020739X.2013.790514

Weber, K., & Mejía-Ramos, J. P. (2011). Why and how mathematicians read proofs: an exploratory study. *Educational Studies in Mathematics, 76*(3), 329-344. doi:10.1007/s10649-010-9292-z

Weber, K., Mejía-Ramos, J. P., Inglis, M., & Alcock, L. (2013). On Mathematicians' Proof Skimming: A Reply to Inglis and Alcock / Skimming: A Response to Weber and Mejía-Ramos. *Journal for Research in Mathematics Education, 44*(2), 464-475.

Wecker, C., Vogel, F., & Hetmanek, A. (2017). Visionär und imposant–aber auch belastbar? Eine Kritik der Methodik von Hatties „Visible Learning “. *Zeitschrift für Erziehungswissenschaft, 20*(1), 21-40.

Weinert, F. E. (1999). *Concepts of competence*. Munich, Germany: Max-Planck-Institut für Psychologische Forschung.

Weiss, M., Herbst, P., & Chen, C. (2009). Teachers' perspectives on "authentic mathematics" and the two-column proof form. *Educational Studies in Mathematics, 70*(3), 275-293. doi:10.1007/s10649-008-9144-2

Wenger, E. (1999). *Communities of practice: Learning, meaning, and identity*. Cambridge, UK: Cambridge university press.

Wickens, C. D., Hutchins, S., Carolan, T., & Cumming, J. (2013). Effectiveness of Part-Task Training and Increasing-Difficulty Training Strategies: A Meta-Analysis Approach. *Human Factors, 55*(2), 461-470. doi:10.1177/0018720812451994

Winter, H. (1983). Zur Problematik des Beweisbedürfnisses. *Journal für Mathematik-Didaktik, 4*(1), 59-95. doi:10.1007/BF03339229

Wittmann, E. C., & Müller, G. (1988). Wann ist ein Beweis ein Beweis. In P. Bender (Ed.), *Mathematikdidaktik: Theorie und Praxis: Festschrift für Heinrich Winter* (pp. 237-258). Berlin, Germany: Cornelsen.

Yackel, E., & Cobb, P. (1996). Sociomathematical norms, argumentation, and autonomy in mathematics. *Journal for Research in Mathematics Education, 27*(4), 458-477.

Yang, K.-L. (2012). Structures of cognitive and metacognitive reading strategy use for reading comprehension of geometry proof. *Educational Studies in Mathematics, 80*(3), 307-326. doi:10.1007/s10649-011-9350-1

Yang, K.-L., & Lin, F.-L. (2008). A model of reading comprehension of geometry proof. *Educational Studies in Mathematics, 67*(1), 59-76.

Zaslavsky, O., Nickerson, S. D., Stylianides, A. J., Kidron, I., & Winicki-Landman, G. (2012). The Need for Proof and Proving: Mathematical and Pedagogical Perspectives. In G. Hanna & M. de Villiers (Eds.), *Proof and Proving in Mathematics Education: The 19th ICMI Study* (pp. 215-229). Dordrecht, Netherlands: Springer Netherlands.

Zazkis, D., Weber, K., & Mejía-Ramos, J. P. (2014). Activities That Mathematics Majors Use To Bridge The Gap Between Informal Arguments And Proofs. In P. Liljedahl, C. Nicol, S. Oesterle, & D. Allan (Eds.), *Proceedings of the 38th Conference of the International Group for the Psychology of Mathematics Education and the 36th Conference of the North American Chapter of the Psychology of Mathematics Education* (Vol. 5, pp. 417-424): Vancouver, Canada: PME.

Zimmerman, C. (2000). The Development of Scientific Reasoning Skills. *Developmental Review, 20*(1), 99-149. doi:10.1006/drev.1999.0497

Zimmerman, C. (2007). The development of scientific thinking skills in elementary and middle school. *Developmental Review, 27*(2), 172-223. doi:10.1016/j.dr.2006.12.001

Zohar, A., & Peled, B. (2008). The effects of explicit teaching of metastrategic knowledge on low- and high-achieving students. *Learning and Instruction, 18*(4), 337-353. doi:10.1016/j.learninstruc.2007.07.001

Zuur, A. F., Ieno, E. N., & Saveliev, A. A. (2009). *Mixed Effects Models and Extensions in Ecology with R*. New York, NY: Springer US.

8 Academic Integrity Statement

I, *Daniel Sommerhoff*, hereby confirm that I completed this Ph.D. thesis with the title:

The Individual Cognitive Resources Underlying Students' Mathematical Argumentation and Proof Skills
From Theory to Intervention

independently, that I have not heretofore presented this paper to another department or university, and I have listed all references used, and have given credit to all additional sources of assistance.

Unterföhring, 4.5.2017

Location & Date

Daniel Sommerhoff

Signature