Mathematical conjecturing and proving: The structure and effects of process characteristics from an individual and social-discursive perspective

Dissertation

an der Fakultät für Mathematik, Informatik und Statistik

der Ludwig-Maximilians-Universität München

eingereicht von

Sarah Ottinger

22.11.2018

1. Gutachter: Prof. Dr. Stefan Ufer

2. Gutachter: Prof. Dr. Ingo Kollar

3. Gutachter: Prof. Dr. Aiso Heinze

Tag der mündlichen Prüfung: 01.02.2019

Danksagung

Ich möchte mich bei all den Menschen bedanken, die zur Entstehung dieser Dissertation beigetragen haben.

An erster Stelle gilt mein Dank meinem Doktorvater Prof. Dr. Stefan Ufer, der diese Arbeit ermöglicht und mit vollster Sorgfalt und Hilfsbereitschaft begleitet hat. Danke für die vielen Möglichkeiten und Einblicke, die du mir gewährt hast - Danke für die großartige Betreuung! Ich konnte viel von dir lernen.

Bei Prof. Dr. Ingo Kollar möchte ich mich ebenfalls für die sehr gute Betreuung und die vielen hilfreichen Anregungen bedanken.

Dank gilt auch Prof. Dr. Aiso Heinze für die Bereitschaft die Drittbegutachtung zu übernehmen, für die zahlreichen Rückmeldungen zu Vorträgen und die bereichernde Zeit in Taipei, im Rahmen des „TaiGer Workshops“.

Bei Prof. Dr. Markus Heydenreich bedanke ich mich für den Prüfungsvorsitz bei meiner Disputation.

Ganz besonderes möchte ich mich bei den (ehemaligen) Kolleginnen und Kollegen am Lehrstuhl der Didaktik der Mathematik der LMU bedanken, die die Zeit der Promotion so unvergesslich schöngemacht haben. Danke für die vielen anregenden, humorvollen und auch tiefgehenden Gespräche. Ein herzliches Dankeschön an Alex, Simon, Alex, Luzia, Anastasia, Timo, Bernhard, Daniela, Tanja, Kati, Ulli, Dani, Hr. Hammer, Lisa, Kathrin, Patricia, Larissa, und ganz besonders an Bine, Laura, Kathrin, Constanze und Daniel.

Daniel danke ich für die schönen gemeinsamen Stunden im Büro, auf Konferenzen und darüber hinaus. Bine danke ich für die vielen intensiven Gespräche und ihr Vertrauen. Kathrin danke ich für den Rückhalt. Constanze danke ich für die gemeinsame Zeit im Büro und die schönen Spaziergänge zum Bahnhof. Laura danke ich für ihre herzliche, unkomplizierte Art und die gemeinsamen Stunden bei einer Tasse Tee.

Dankeschön auch an Frau Prof. Dr. Gasteiger, die zu einer so wundervollen Zusammenarbeit am Lehrstuhl beigetragen hat.

Bei Dr. Matthew Inglis möchte ich mich für den erkenntnisreichen Forschungsaufenthalt an der University of Loughborough bedanken.

Mein Dank geht auch an das gesamte „ELK“-Team: Dr. Freydis Vogel, Dr. Matthias Schwaighofer, Anselm Strohmaier, Ilka Terwedow, Prof. Dr. Stefan Ufer, Prof. Dr. Ingo Kollar, Prof. Dr. Kristina Reiss und Prof. Dr. Frank Fischer. Diese Dissertation ist im Rahmen der ELK-Math-Studie “Effects of heuristic worked examples and collaboration scripts on the acquisition of individual and discoursive components of mathematical argumentation skill“, gefördert von der DFG, entstanden.

Ich denke oft an Matthias, der auf so tragische Weise aus den Leben gerissen wurde. Seine persönliche Art und seine Fachkenntnisse habe ich sehr geschätzt.

Weiterhin möchte ich mich bei dem Mathematikdidaktik-Team der TU München unter Leitung von Frau Prof. Dr. Kristina Reiss für den gemeinsamen Austausch im Rahmen des Oberseminars bedanken.

Vielen Dank auch an die Professoren und Doktoranden der ersten und zweiten Kohorte des Doktorandenkollegs „Scientific Reasoning and Argumentation“, gefördert vom Elitenetzwerk Bayern.

Ein ganz großer Dank geht an meine Eltern - Angelika und Franz - die mich auf meinem Weg begleitet und bedingungslos unterstützt haben. Es ist so wertvoll zu wissen, dass ich mich immer auf euch verlassen kann.

Danke auch an meine Schwester Melanie und ihren Freund Béla für die emotionale Unterstützung und euren Rückhalt.

Der letzte Dank geht an meine eigene kleine Familie – Tobi und Oskar. Ich bin so unglaublich dankbar darüber, dass es euch gibt. Ihr gebt mir Kraft und Halt – ihr zeigt mir wie schön das Leben sein kann. Danke dafür.

Summary

Most university courses in mathematics programs are characterized by a strong focus on the axiomatic nature of mathematics, and thus also on proof as the central scientific method of mathematics (Selden, A. & Selden, 2008). Lecturers write proofs on the blackboard, students attempt to demonstrate their understanding and skills by proving theorems on their own or in collaboration with others. However, there is often little systematic discussion in these courses on how new mathematical conjectures can be generated and on how proofs are constructed (Alcock, 2010). Students' experiences with conjecturing and proving in schools or in university mathematics courses often lead them to "consider proof as a static product rather than a negotiated process that can help students justify and make sense of mathematical ideas" (Otten, Bleiler-Baxter, & Engledowl, 2017, p. 112). Yet, several authors (e.g., Epp, 2003; Savic, 2015a; Selden, A. & Selden, 2008) have hypothesized that often only little time can be devoted to illustrate students which strategies and processes may help to step through the proof construction process and to recover from proving impasses. Furthermore, the knowledge about what characterizes proof processes that lead to a successful outcome (i.e., an acceptable mathematical proof [according to local acceptance criteria]) is rare.

To approach this issue, an extensive systematic literature search was conducted to summarize common claims and empirical findings about promising conjecturing and proving processes. 126 articles that focussed on conjecturing and proving were clustered using a topic modeling method. The algorithm identified 17 different topics. The most representative papers for each topic, in total 45 papers, were qualitatively analysed with regard to their research perspectives on which they were based and their claims and findings about the processes that are needed to successfully generate conjectures and construct proofs. This combination of statistical clustering and qualitative analyses allowed a systematic categorization of claims and empirical findings about successful conjecturing and proving processes in the literature. Based on this review, a set of characteristics of conjecturing and proving processes, that are assumed or reported to be crucial for success, is proposed.

For the further analysis of such process characteristics, we started from a model differentiating students' prerequisites they bring to bear on the proving situation, the conjecturing and proving processes they engage in, and the quality of the resulting product. The main question of the empirical work in this dissertation was, which process characteristics influence the quality of the final product (the formulated conjecture and constructed proof), and in which way they mediate the impact of students' prerequisites on this product. Specifically, we distinguished between individual-mathematical and social-discursive process characteristics of conjecturing and proving. These process characteristics were extracted from prior research in mathematics education or in educational psychology or in the Learning Sciences.

The central aim of this dissertation was to develop an instrument for assessing (prospective undergraduate) mathematics students' conjecturing and proving processes in collaborative situations. A high-inference rating scheme with seven scales, based on theoretical considerations and on rating guidelines adapted from educational research was designed. The rating scheme was evaluated in a study with *N=98* prospective undergraduate students working in dyads on an open-ended conjecturing and proving task. The results of the empirical study with regard to the basic analyses showed that collaborative conjecturing and proving processes could be rated with sufficient reliability and that the structure of the data corresponded to the underlying theoretical assumption that two dimensions, one related to individual-mathematical and one related to social-discursive process characteristics can be distinguished. The in-depth analyses pointed out that individual-mathematical process characteristics were predictive for the quality of the resulting product and mediated the relation between prerequisites (students' prior knowledge on proof) and the quality of the product.

In this way, the dissertation contributes to the scientific debate on how to assess (mathematical argumentation) skills (e.g., Blömeke, Gustafsson, & Shavelson, 2015; Koeppen, Hartig, Klieme, & Leutner, 2008) and provides theoretical and empirical insights on individual-mathematical and social-discursive process characteristics that describe the quality of collaborative conjecturing and proving processes.

Zusammenfassung

Die meisten universitären Mathematikveranstaltungen zeichnen sich durch einen starken Fokus auf den axiomatischen Charakter der Mathematik und damit auch auf das Beweisen als zentrale wissenschaftliche Methode der Mathematik aus (Selden, A. & Selden, 2008). Die Dozierenden schreiben Beweise an die Tafel, die Studierenden versuchen, ihr Verständnis und ihre Fähigkeiten darzulegen, indem sie Theoreme alleine oder in Zusammenarbeit mit anderen beweisen. In diesen Veranstaltungen wird jedoch häufig wenig systematisch diskutiert, wie neue mathematische Vermutungen gefunden und wie Beweise konstruiert werden können (Alcock, 2010). Die Erfahrungen der Studierenden mit Vermutungen und Beweisen, die sie während ihrer Schulzeit oder auch innerhalb der universitären Veranstaltungen gesammelt haben, führen sie oft dazu, *den Beweis als statisches Produkt zu betrachten und nicht als ausgehandelten Prozess, mit dessen Hilfe mathematische Ideen begründet und verstanden werden können* (Otten, Bleiler-Baxter & Engledowl, 2017, S. 112). Mehrere Autoren (u.a., Epp, 2003; Savic, 2015; Selden, A. & Selden, 2008) haben jedoch die Hypothese aufgestellt, dass oft nur wenig Zeit aufgewendet werden kann, um den Lernenden zu zeigen, welche Strategien und Prozesse dabei helfen können, Beweise zu generieren und Fehlwege zu überwinden. Außerdem fehlt es noch an belastbarem Wissen darüber, was Beweisprozesse charakterisiert, die zu einem erfolgreichen Ergebnis führen (d.h. zu einem akzeptablen mathematischen Beweis [gemäß den lokalen Akzeptanzkriterien]).

Um diese Problematik anzugehen, wurde eine umfangreiche systematische Literaturrecherche durchgeführt, die die weit verbreiteten Behauptungen und empirischen Befunde zu vielversprechenden Conjecturing- und Beweisprozessen zusammenfasst. 126 Artikel, die sich auf Vermutungen und Beweise fokussieren, wurden mithilfe einer „Themenmodellierungsmethode" zu einzelnen Themensträngen geclustert. Mithilfe des Algorithmus konnten 17 verschiedene Themenstränge identifiziert werden. Die für jeden Themenstrang repräsentativsten Artikel, insgesamt 45 Artikel, wurden hinsichtlich ihrer Forschungsperspektiven und ihrer Behauptungen und Erkenntnisse über die Prozesse, die zur erfolgreichen Formulierung von Vermutungen und zur Generierung von Beweisen erforderlich sind, qualitativ analysiert. Diese Kombination aus statistischem *Clustering* und qualitativen Analysen ermöglichte eine systematische Kategorisierung von Behauptungen und empirischen Befunden über erfolgreiche Conjecturing- und Beweisprozesse in der Literatur. Basierend auf dieser Kategorisierung wird eine Reihe von Merkmalen von Conjecturing- und Beweisprozessen präsentiert, von denen angenommen oder berichtet wird, dass sie für den Erfolg entscheidend sind.

Grundlage für die weitere Analyse derartiger Prozessmerkmale stellte ein Modell dar, welches zwischen den individuellen Voraussetzungen, den Conjecturing- und Beweisprozessen sowie

der Qualität des daraus resultierenden Produkts unterscheidet. Dem empirischen Teil dieser Dissertation liegt die zentrale Frage zugrunde, welche Prozessmerkmale prädiktiv für die Qualität des Beweisproduktes sind (die Qualität der formulierten Vermutung und des konstruierten Beweises) und inwiefern diese Prozessmerkmale den Einfluss der individuellen Voraussetzungen auf die Qualität des Produktes mediieren. Insbesondere wurde in diesem Projekt zwischen individuell-mathematischen und sozial-diskursiven Prozessmerkmalen des Conjecturings und Beweisens unterschieden. Diese Prozessmerkmale wurden aus früheren Forschungsarbeiten aus dem Bereich der Mathematikdidaktik, der Psychologie oder den Learning Sciences abgeleitet.

Zentrales Ziel dieser Dissertation war die Entwicklung eines Analyseinstruments zur Beurteilung der kooperativen Conjecturing- und Beweisprozesse von (zukünftigen) Mathematikstudierenden in Hinblick auf individuell-mathematische und sozial-diskursive Prozessmerkmale.

Es wurde ein hoch-inferentes Bewertungsschema mit sieben Ratingskalen entwickelt, das auf theoretischen Überlegungen und Bewertungsrichtlinien basiert, die aus der Bildungsforschung abgeleitet und adaptiert wurden. Das Bewertungsschema wurde im Rahmen einer Studie mit $N = 98$ Studienanfänger/- innen, die in Dyaden an einer offenen Conjecturing- und Beweisaufgabe arbeiteten, eingesetzt. Die Ergebnisse der empirischen Studie im Hinblick auf die Basisanalysen zeigten, dass kooperative Conjecturing- und Beweisprozesse hinreichend zuverlässig bewertet werden können und dass die Struktur der Daten der zugrundeliegenden theoretischen Annahme entsprach, dass zwei Dimensionen, eine die sich auf die individuell-mathematischen Prozessmerkmale und eine die sich auf die sozial-diskursiven Prozessmerkmale bezieht, unterschieden werden können. Die weiteren Analysen zeigten auf, dass die individuell-mathematischen Prozessmerkmale für die Qualität des resultierenden Produkts prädiktiv waren und die Beziehung zwischen den Voraussetzungen (dem Vorwissen der Studierenden über Beweise) und der Qualität des Produkts mediiert haben.

Auf diese Weise trägt die Dissertation zur wissenschaftlichen Debatte, wie (mathematische Argumentations-) Kompetenzen beurteilt werden können, bei (u.a., Blömeke, Gustafsson & Shavelson, 2015; Koeppen, Hartig, Klieme & Leutner, 2008) und liefert theoretische und empirische Einblicke zu individuell-mathematischen und sozial-diskursiven Prozessmerkmalen, die die Qualität von kooperativen Conjecturing- und Beweisprozessen beschreiben.

Table of Contents

1 General Introduction

In the secondary and especially in the tertiary education, conjecturing and proving as specific types of mathematical argumentation are central activities. Producing a conjecture and an acceptable mathematical justification for it has shifted into the focus of mathematics curricula worldwide (e.g., Common Core State Standards Initiative, 2010). However, the ability to construct arguments for or against mathematical claims and to formulate new conjectures are a challenging demand for students at all educational levels. Much of the research on proof construction in the context of mathematics education has been concerned with difficulties students encounter (e.g., Epp, 2009; Moore, 1994; Selden, A. & Selden, 2008) or cognitive resources that are found to affect students' proving performance (e.g., Ufer, Heinze, & Reiss, 2008; Weber, 2001). Understanding how to develop the key insights that are needed to construct a proof (e.g., Raman, 2003) and how to write them down (e.g., Selden, A. & Selden, 2009) have always been considered as important goals by mathematics educational researchers and teachers (Stylianides, G. J., Stylianides, & Weber, 2017).

There is a general interest in studying the processes involved in learning and problem solving in advanced mathematics (Dreyfus, 2002) and, in particular, examining the complex cognitive processes related to proof construction (Cai, Mamona-Downs, & Weber, 2005). One reason for that is to generate theoretical knowledge about the mental and physical processes high-achieving students or mathematicians engage in during proof construction, and to understand why they are employing a specific strategy or process (Selden, A., McKee, & Selden, 2010).

Some practical issues also guide this strand of research: For instance, teachers of advanced mathematics courses should become more conscious of which process characteristics determine the success of conjecturing and proving, and explicitly introduce those processes that appear to be relevant in their teaching (cf. Selden, J. & Selden, 2015). Furthermore, scaffolds to foster mathematical argumentation skills such as heuristic worked examples or collaboration scripts that have been developed in the past are mainly based on expert-models (e.g., Kollar et al., 2014; Schwaighofer et al., 2017). Yet, before creating such learning environments and introducing process characteristics in lectures, it might be essential to find out where novice students actually require support in the first place, and which of their processes have to be encouraged most. Several researchers claim that the processes used by novices differ from those used by experts (e.g., Nadolski, Kirschner, & van Merriënboer, 2006). From this point of view, it is important to figure out what the components of successful conjecturing and proving processes of novice students are and how they are related to each other.

In this dissertation, we reviewed literature with regard to the process characteristics that are considered to be relevant for success from a theoretical perspective (study I – research review) and analysed the conjecturing and proving processes of novice students (study II & study III – empirical studies). Moreover, we were interested in studying to what extent these process characteristics depend on students' prerequisites.

It is widely agreed that conjecturing and proving are complex skills including processes such as generating hypotheses from examples (e.g., Ellis et al., 2017; Koedinger, 1998; Philipp, 2012) or finding a chain of claims that can be worked out to a deductive proof (e.g., Boero, 1999; Stylianides, A. J., 2007). These processes can be described as individual-mathematical ones. Since mathematical knowledge generation is often embedded in social contexts (e.g., seminars, classrooms, or small collaborative settings), participating successfully in mathematical debates is vital for conjecturing and proving as well (Vidakovic & Martin, 2004). Thus, conjecturing and proving skills comprise individual-mathematical and social-discursive process characteristics (Kollar et al., 2014). Even though critical elements of substantial collaborative argumentation processes have been put forward in several domains (e.g., Chi & Wylie, 2014; Weinberger & Fischer, 2006) and in mathematics education (e.g., Mueller, Yankelewitz, & Maher, 2012; Pease, A. & Martin, 2012), there is still limited knowledge about the structural relationship between individual-mathematical and social-discursive process characteristics. Moreover, empirical findings about which process characteristics positively influence the quality of the resulting product (i.e., the produced conjecture and proof) on the one hand and about their relation to prior knowledge on proof on the other hand, are rare.

The central aim of this dissertation project was to develop an instrument to describe the process quality of collaborative mathematical argumentation and proving and to use this instrument to identify key characteristics of successful conjecturing and proving processes in collaborative situations.

To form a solid theoretical base for this dissertation, Chapter 2 summarizes the state of research on conjecturing and proving in the context of the transition from secondary to tertiary mathematics education, and defines how central concepts and terms including *argumentation*, *conjecturing*, and *proving* are understood here in this thesis. Section 2.4 presents the central model that serves as a basis for our empirical research, which distinguishes between *prerequisites, processes*, and *performance* (respectively the *quality of the final product*). In Chapter 3, we discuss the motivations behind the identification of individual-mathematical and social-discursive process characteristics. The resulting overall research questions and aims of this dissertation project are described in Chapter 4.

Chapter 5 includes a literature review. The purpose of this review is (i) to organize and analyse past research on conjecturing and proving (processes) and (ii) to categorize common claims and empirical findings about successful conjecturing and proving processes. Using a topic modeling method, which allows to cluster a collection of documents by implementing a statistical algorithm (for discovering the latent semantic structures within these documents), we identified 17 topics within the literature on conjecturing and proving. Results indicate that the literature on conjecturing and proving covers topics that are related to the proving as problem-solving, proving as convincing, and proving as a socially-embedded activity perspective (cf. Stylianides, G. J. et al., 2017) as well as topics that refer to the discovery perspective on proof or on automatic theorem proving. Categories of successful conjecturing and proving processes are extracted from studies' reports on how successful mathematicians, university students, or high-achieving college students employ specific proving processes and from claims about which proving processes are assumed to be crucial for the success. We distinguish between categories of sub-goals within conjecturing and proving processes necessary for success and categories of process characteristics that are assumed to be helpful in achieving these sub-goals. The analysis yields a broad range of categories of sub-goals such as *developing a strong understanding of the statement to be proved/ estimation of the truth, resolving fixations*, or *translating less formal to formal arguments* as well as categories of process characteristics such as *varying examples systematically, considering boundary cases, using formal symbols and algebraic representations*, or *applying the trial-and-error strategy*.

Chapter 6 provides an extensive description of the development of the high-inference rating scheme that was designed to assess (undergraduate) mathematics students' collaborative conjecturing and proving processes. In this chapter, it is reported how process characteristics were deduced from the literature and operationalized, how the quality levels of each rating scale were defined, and how the rater training was structured and conducted.

Chapter 7 and Chapter 8 comprise two analyses of one empirical study that is based on the data of *N=98* prospective university mathematics students participating in a voluntary preparatory course. For the analyses, a set of 49 recordings of dyadic collaboration processes is used, taking verbal and written contributions of all participants into account.

The first analysis (presented in Chapter 7) investigates the empirical structure of individual-mathematical and social-discursive process characteristics, which were inferred from the literature. Results indicate that collaborative conjecturing and proving processes can be described by a two-dimensional construct, comprising either mostly individual-mathematical or mostly social-discursive process characteristics.

The second analysis (described in Chapter 8) focuses on the predictive power of individual-mathematical and social-discursive process characteristics. The theory-based process characteristics were validated against the quality of the produced conjectures and proofs as the resulting outcome of students' collaborative conjecturing and proving processes. Furthermore, the students' prior knowledge on proof has been taken into account. Results show that the individual-mathematical component of collaborative conjecturing and proving processes is predictive for the quality of the resulting outcome. Especially, generating accurate and structurally sound arguments during the collaborative proving discourse can be considered as key characteristics of successful collaborative conjecturing and proving processes. At the end of this dissertation, a summary of the main findings is given.

In Chapter 9, limitations and implications for research and teaching are discussed.

2 State of research

2.1 The transition phase

The empirical study of this dissertation is situated in the transition phase from secondary to tertiary mathematics education. Before reviewing definitions and models of mathematical argumentation, conjecturing and proving and formulating the specific research questions of this dissertation, the following three sections present how the character of the learning domain mathematics changes at the transition from secondary school to university mathematics, how the focus on formal concepts and proofs increases, and what students' difficulties with proof construction are.

2.1.1 Teaching and learning of mathematics at school and at university – challenges at the transition

Mathematics educational researchers have publishing work that deals with mathematics learning and teaching at the university level with a specific focus on the challenges at the transition phase from school to university for more than 20 years (Artigue, 2001). In particular, the high drop-out rates in mathematics-related academic study programs (e.g., Dieter, 2012) urged researchers to pay more attention to the discrepancies between the two educational systems and the resulting difficulties students encounter (e.g., Artigue, 2001; Heublein, 2014; Kosiol, Rach, & Ufer). It has been reported that the different teaching styles, learning contexts, and assessment strategies contribute to the transitional gap between the secondary and academic sector (Thomas, M. O. J. & Klymchuk, 2012).

Thomas and Klymchuk (2012) have hypothesized that the large course sizes at university might be one factor that leads to less interaction and communication with students. In comparison to school, the teaching style at university is more teacher-centred and the time to "'do' problems" (p. 289) is limited. It can be said that it is more difficult for students to ask individual questions (Thomas, M. O. J. & Klymchuk, 2012) and that the amount of personal attention students get from their teachers decreases extremely compared to the school context (Gruenwald, Klymchuk, & Jovanoski, 2004).

In general, the academic field is characterized by a high-degree of self-learning phases. University students have to take great responsibility for their own learning, acquire self-regulative techniques as well as elaboration strategies (Rach, 2014; Rach & Heinze, 2011). The type of problems students have to deal with changes as well. Secondary school tasks are frequently split into simpler sub-tasks and provide hints that may encourage students to develop a solution (Praslon, 2000). Furthermore, most of the tasks students have to work on are routine problems that are analogous to those already demonstrated by the teacher and that do not involve any conceptual obstacles, so that they can be described as tasks with a low cognitive potential (Jordan et al., 2008). Such problems constitute exercises and might not be regarded as problem-solving tasks (Selden, A. & Selden, 2013b). At university, applying routines is usually not sufficient as the problems that students face are more complex (Gruenwald et al., 2004).

The survey of de Guzmán, Hodgson, Robert, and Villani (1998) pointed out that a large proportion of university students regret that the lectures do not follow a particular textbook and that concrete examples are rarely given. From students' perspective, the teaching style at the university is often too abstract. They miss backboard drawings or hand-out notes with detailed explanations. Informal content is presented as well, but mainly only orally. It could be observed that students typically copy written content, but not necessarily oral comments in their notes (Fukawa-Connelly, Weber, & Mejía-Ramos, 2017; Weber, Fukawa-Connelly, Mejía-Ramos, & Lew, 2016)

Regarding the assessment culture at school, the research survey by Thomas and Klymchuk (2012) has demonstrated that there are a lot of internal (and external) assessments and much more emphasis on passing the exams than on learning to understand. They reported that some school teachers have the impression that they only teach to assessments. Furthermore, the results of their survey demonstrated that most students prefer the assessment methods at the tertiary level for several reasons such as questions are more to the point and without hierarchical style, and also the precision of solutions is rewarded. Yet, as reported in Gruenwald et al. (2004), university lecturers see the necessity for changing the assessment style at university as well. Some of them suggest to demand weekly tests and oral exams as both methods may allow to give students a more detailed feedback than just written exams at the end of the semester. Kahn and Hoyles (1997) claimed that there has already been made a change towards more continuous assessments.

Based on the assumption that students and expert mathematicians view advanced mathematics from different perspectives, Weber et al. (2016) suggested that it has to be clearly communicated to students what they should know and learn from the lectures they attend. The results of the study of Gruenwald et al. (2004) supported that it is not always obvious to the students what is expected of them. More attention needs to be paid to the communication on how to achieve the goals required to pass academic mathematics courses. Even at school, rules and social-mathematical norms are often not clearly discussed. Thus, students have problems to decide whether a proof is valid or not, or what is accepted as explanation. They do not feel responsible for these aspects of mathematics (e.g., Dreyfus, 2002; Gueudet, 2008).

University students have to take more responsibility for their own learning (e.g., Rach & Heinze, 2011) and learn to recover quickly from failures and disappointments (Selden, A. & Selden, 2013a). Schiefele, Streblow, and Brinkmann (2007) concluded that personal traits such as intrinsic learning motivation, self-confidence, persistency, social skills, and the ability to cope with performance-related pressure also need to be taken into account when focusing on the discontinuity phenomena of mathematics students during their transition from secondary to tertiary education.

The most relevant aspect contributing to the transitional gap is that faculty members of academic mathematics courses place high value on formal concepts, accuracy, and deductive proofs (e.g., Thomas, M. O. J. & Klymchuk, 2012). These features characterize mathematics as a scientific discipline. In the following section, it is described how mathematics is presented at school, and how mathematics is taught at university from a scientific perspective with a specific emphasize on proofs and formal representations.

2.1.2 The character of mathematics at school and at university and the role of proof

Epp (2003) and Gueudet (2008) described what happens during the secondary-tertiary transition by using the metaphor that novice students often feel like a foreigner entering a new world, or at least a new country in which a different language is spoken and other laws are effective. These are the language and rules that mathematicians use to construct and communicate proofs.

In the secondary school, there is a specific focus on technical aspects such as manipulating algebraic expressions, calculating derivatives, and applying formulas. Concepts and procedures are considered as tools for describing more or less real life situations and solving everyday problems (e.g., Vollstedt, Heinze, Gojdka, & Rach, 2014). According to this, topics such as fractions, percentages and area calculations, which are rarely of interest from a scientific point of view, have a high priority in the school syllabus (Rach & Heinze, 2011). School mathematics can be described as very mechanical and situational (Gruenwald et al., 2004). Some researchers assumed that students often succeed in school mathematics by employing an algorithm without understanding the concepts beyond (e.g., Guzmán et al., 1998; Tall, 1991). Furthermore, some authors from university mathematics education have critically remarked that the teaching style at school encourages students to learn disjointed facts and procedures, and push the theory into the background (Gruenwald et al., 2004). Aspects that characterize mathematics as a scientific discipline (e.g. building a coherent and consistent theory, deductive proofs, and formal definitions) are rather underrepresented, even in a propaedeutic form, and only sporadically implemented in the school curriculum (Rach & Heinze, 2011). Therefore, students may experience substantial difficulties, when entering the tertiary level (Guzmán et al., 1998). The gap between school and university mathematics can be considered as a great leap from empirical to abstract mathematics, from less formal to formal representations (Nardi, 1996). Students have to learn an entirely new way of thinking (Tall, 1991). Moore (1994) claimed that students are inadequately prepared for the rigor and accuracy that is expected from them at the university. Some authors hypothesized that students have little idea of what mathematics is, when entering academic mathematics courses, and that they take the view that it is solely an extension of school mathematics (Hoyles, Newman, & Noss, 2001; Nardi, 1996).

The teaching content at the university is organized and demonstrated in a specific axiomatic and rigorous way, and comprises formally defined abstract concepts, theorems, logical deductions, and proofs. The lectures follow a specific consistent shape, the so called DTP (Definition-Theorem-Proof) structure (Engelbrecht, 2010; Hoyles et al., 2001). Therefore, proofs achieve a new and important status at the tertiary level (Guzmán et al., 1998).

In the study of Harel and Sowder (1998), university students' proving attempts have been categorized. They identified that most students evaluate the validity of a proof by referring to external factors or an external authority. Analytical proof schemes, where conviction relies on logical deduction were rarely observed. It has been shown that only the minority of mathematics students are able to construct a coherent chain of arguments that is accepted as proof by the mathematical community (e.g., Gueudet, 2008; Moore, 1994). Proofs at the tertiary level tend to be more complex, and have to be based on formal definitions and previously established theorems (Selden, A. & Selden, 2009). Students have to develop a deep and conscious knowledge of the logical principles involved, and being able to employ them (Epp, 2003). Informal and empirical arguments such as examples, which are often accepted as justification for a statement in the context of school mathematics, can still be used to explore the problem situation initially and to think things through, but finally such intuitive or informal reasoning must be made more formal and precise for communication and presenting purposes in the context of university mathematics (Hanna, Jahnke, & Pulte, 2010; Selden, A. & Selden, 2009).

Even though there is a strong emphasis on enhancing students' creativity and informal conceptual understanding (Selden, A. & Selden, 2009), university students are mainly assessed on their ability to produce formal mathematics. Consequently, university students frequently assume that formal aspects are superior to all other aspects of mathematics. For instance, they focus more on using formal symbols than constructing a coherent chain of arguments (Weber et al., 2016).

From the findings reported in the literature and presented in the two last sections, we conclude that the transition from secondary to tertiary mathematics education is a challenging phase for most students. Reasons for that have been attributed to changes in the learning and teaching culture, in the assessment methods, in the character of mathematics taught, and in particular in the role of proofs.

2.1.3 Students' difficulties with proofs

It is well known that many students have difficulties in generating conjectures and constructing proofs. This is hypothesized to be one reason for the high dropout rates, as students' proofs are used as an important component in evaluating and grading their understanding and performance in content courses, such as real analysis or linear algebra (Selden, A. et al., 2010). Students' difficulties related to conjecturing and proving at different educational levels have been identified and discussed by a number of researchers (e.g., Epp, 2009; Koedinger, 1998; Moore, 1994; Selden, A. & Selden, 2008; Selden, A. & Selden, 2011). In the following section, we summarize the main findings about students' conjecturing and proving difficulties that have been observed and documented in the literature.

Several researchers reported that undergraduate students tend to focus more on procedures than on content, more on formal aspects than on understanding the concepts involved (e.g., Moore, 1994). This could be a result of the teaching-style they have experienced at the secondary level. Memorizing and imitating proofs may lead students to face problems, not only with producing proofs, but even with recognizing what a proof is (Chazan, 1993; Moore, 1994; Raman, 2003). It has been observed that students often fail to discover, interpret, or use theorems on their own (Selden, A. & Selden, 2008; Weber, 2001). Some students already struggle with reviewing their notes if there are any relevant lemmas or theorems they could apply. They are prone to proceed directly from the definitions involved and write the entire definition into a proof, rather than just saying that the definition applies to a particular mathematical object. The distinction between using a definition and examining whether an object satisfies a definition remains unclear for them. These difficulties influence students' ability to handle the problem-solving aspects of proofs (Selden, A. & Selden, 2011).

Furthermore, students often do not know how to begin and end, for instance, direct or contradiction proofs (Moore, 1994; Selden, A. & Selden, 2011), or what to do next (Selden, A. et al., 2010). They have problems with applying standard proving techniques and with unpacking the logical structure of (informally stated) theorems. These are considered as some of the reasons why they are not able to structure proofs (e.g., Selden, A. & Selden, 2011).

A further category of proving difficulties involves that some students do not recognize the constraints of empirical or authoritative evidence (Stylianides, G. J. & Stylianides, 2009). It is claimed that mathematics researcher, lecturers and students have disparate views on mathematics and thus, their conceptions of what constitutes evidence and justification in mathematics may differ (e.g., Thomas, M. O. J. & Klymchuk, 2012). The study of Martin and

Harel (1989) has shown that many pre-service teachers accepted empirical or even incorrect arguments as proofs. Coe and Ruthven (1994) investigated the proof practices and constructs of advanced mathematics students that followed a reform-based curriculum and also found out that students predominantly prefer empirical proof strategies.

Transforming informal into formal arguments represents a further challenge (Zazkis, Weber, & Mejía-Ramos, 2016). Students' difficulties in using formal-symbolic notations and the specific mathematical language are well documented in the literature. These include problems such as taking the scope (Epp, 2003) and order of quantifiers into account (Dubinsky & Yiparaki, 2000), understanding that the value of a variable can be arbitrary, but fixed and does not change its value within one algebraic expression (Epp, 2003).

Besides these mathematical difficulties, Selden, A. and Selden (2011) also cited some difficulties, such as the incorrect copying of a definition from the blackboard or the incorrect articulation of notations and terms when reading or explaining a proof in one's one words, which they summarized as "non-mathematical proving difficulties" (p. 678).

In general, the descriptions of students' difficulties provide insights at the process level (though sometimes derived from students' written proof attempts). Unfavourable sub-processes or sub-processes that are often not handled correctly have been discussed. Yet, would it be more promising to look at the processes that actually lead to success or that make the difference between successful and less successful proving processes? In this dissertation, we will focus on and investigate the process characteristics that are assumed to be crucial for success. Furthermore, as the success of proving processes are primarily determined by their outcomes, we take the quality of the resulting product into account as well.

For better understanding the difficulties and obstacles that students face in relation to proof, researchers have begun to search for the origins and sources of such difficulties (Mariotti, 2006). In Chapter 2.2.3, we will present the cognitive and affect-motivational resources that have been considered as prerequisites that may have an important influence on the proving performance (on the quality of the resulting proof product).

2.2 Argumentation, conjecturing and proving

In this chapter, the concepts of conjecturing and proving are described from an individual and social-discursive perspective. We consider the characterizations of argumentation, conjecturing and proving in mathematics and their relationships (section 2.2.1). We provide an overview of the context- and personal-specific factors that have been discussed in the literature (section 2.2.2 and section 2.2.3) and introduce different approaches to conceptualize mathematical argumentation skills (section 2.2.4). Different models of argumentation and the proof construction process (section 2.2.5) are presented and compared in terms of what we already know about conjecturing and proving processes. We derive gaps and open questions from the current state of research on conjecturing and proving. Since conjecturing and proving are often embedded in social contexts, we also refer to the social-discursive perspective on mathematical argumentation and discuss the role of peer collaboration within conjecturing and proving activities (section 2.2.6). We conclude with a short summary presenting our research model.

2.2.1 Defining key terms

2.2.1.1 Different perspectives on proof

The terms *argumentation, conjecturing,* and *proving* have been used in different ways. Even though several researchers and curriculum frameworks emphasize the importance of argumentation and proving throughout the grades (e.g., Common Core State Standards Initiative, 2010; Hanna, 1995, 2000), there are various views on how proof (the corresponding concept to the activity of proving) can be defined (Mariotti, 2006; Stylianides, A. J., 2007), and how it is related to conjecturing and argumentation (e.g., Pedemonte, 2007). From a mathematical perspective, proofs are associated with formal definitions and theorems (the use of already established mathematical results), and logical deductions that interlink the assumptions (that are regarded as true) with the conclusions (e.g., Healy & Hoyles, 2000). Griffiths (2000) stated that “a mathematical proof is a formal and logical line of reasoning that begins with a set of axioms and moves through logical steps to a conclusion” (p. 2). In mathematics curricula all over the world, (deductive) reasoning is considered as a crucial learning goal and refers “to this family of activities that are frequently involved in the development of proofs: identifying patterns, making conjectures, and providing arguments - both proofs and arguments that do not meet the standard of proof” (Stylianides, G. J., 2010, p. 39). As already noticed by Balacheff (1991), most characterizations of proof point mainly to the logical structure of proofs. Different types of proofs can be distinguished with regard to the underlying logical structure, the proving techniques that have been applied to develop a proof (e.g., proof by exhaustion, proof by mathematical induction, proof by contradiction), and to the type of claims that have to be proven (e.g., existence proofs) (Hanna, Villiers, & International Program Committee, 2008). Besides of establishing the truth or falsity of an assertion (proof as a means of verification/falsification) and organizing results into a deductive system consisting of axioms, concepts and theorems (proof as a means of systematization), proofs can serve a broad range of functions such as providing insight into why an assertion is true or false (proof as a means of explanation) or leading to new results (proof as a means of discovery). Other functions that proofs can fulfil are tackling a (new) intellectual challenge and providing satisfaction afterwards (proof as a means of intellectual challenge) as well as reporting and disseminating mathematical knowledge (proof as a means of communication). In this dissertation, we mainly focus on the functions of proofs related to verification, systematization and communication (Villiers, 1999).

Mariotti (2006) pointed to the cognitive side of mathematics and in particular of proofs, and to their integration into a social context. She emphasized that after a successive phase of (empirically) discovering and systematizing ideas and arguments, a phase follows in which the body of developed knowledge is made accessible to the scientific community. In that sense, the creative phase of discovery and systematization describes the cognitive dimension, and the phase of communication refers to the social side of proof. The quote "...it appears that proof is a form of **discourse**, a means of communication among people doing mathematics" (bold added) (Volmink, 1990; as cited in Villiers, 1999) underlines the importance of the social side of proof. Manin's (1977) statement that “a proof becomes a proof after the social act of accepting it as a proof” (p. 48) is consistent “with the conceptualization of proofs as nonabsolute objects” (Stylianides, A. J., 2007, p. 298).

Having these different conceptualizations of proof in mind, we go along with the definition proposed by Stylianides, A. J. (2007, p. 291), describing proof in the following way:

“*Proof* is a *mathematical argument*, a connected sequence of assertions for or against a mathematical claim, with the following characteristics:

1. It uses statements accepted by the classroom community (*set of accepted statements*) that are true and available without further justification;

2. It employs forms of reasoning (*modes of argumentation*) that are valid and known to, or within the conceptual reach of, the classroom community; and

3. It is communicated with forms of expression (modes of argument representation) that are appropriate and known to, or within the conceptual reach of, the classroom community.”

Even though the definition originally emerged in the context of proofs at the secondary mathematics education, it is also transferable to the context of university mathematics. In this way, the classroom community consists of mathematicians such as professors and lecturers as well as of mathematics students. We conclude that proofs in mathematics, as deductive chains of arguments that are based on true statements, valid forms of reasoning, and appropriate forms of representations, are context-dependent (Thurston, 1998). Furthermore, we interpreted the term *valid* as an expression for the fact that the validity of a proof is usually determined by certain criteria defined by the corresponding mathematical community. Such criteria include *sociomathematical norms* (Yackel & Cobb, 1996) and values (Dawkins &

Weber, 2017). According to Stylianides, G. J. et al. (2017) the activity in search for a proof is called proving.

Yet, what are the common aspects between argumentation and proving, and why is it important to take the differences between them into account? The debate about what constitutes a mathematical proof leads directly to the question of the relationships between conjecturing, proving, and argumentation.

2.2.1.2 The relationships between proving, conjecturing and argumentation

Before clarifying the relationship, we want to present different perspectives on mathematical argumentation.

Delineating the terms argumentation, argument, conjecturing and proving

As with the concept of proof, there is no universally accepted definition of argumentation in mathematics education (Pedemonte, 2007). Argumentation is used for both describing "the process which produces a logically connected (but not necessarily deductive) discourse about a given subject […] and the text produced by that process" (Douek, 2007, p. 169). The process of producing a logically connected discourse comprises phases of identifying reasons, making inductions, drawing conclusions and applying them to the subject the discourse focuses on. An argumentation is a sequence of arguments (including drawings, examples, verbal arguments etc.) and inferences, whereas an argument represents a reason or a structured chain of reasons for or against a statement or opinion (e.g., Douek, 2007; Hornikx & Hahn, 2012; Toulmin, 1958). Argumentations (or arguments) are either produced individually (for instance, in a written form) or produced orally embedded in a social context (Douek, 1999).

Some authors emphasized the discursive character of argumentation (e.g., Alibert & Thomas, 1991). In comparison to proof, argumentation with its non-constraining character (Perelman, 1979) leaves some degree of freedom regarding the type of inferences (inductive, abductive, deductive) chosen (Douek, 1999; Pedemonte, 2008). It usually takes place informally and incidentally within mathematicians to refine, discuss or communicate mathematical problems and outcomes. Some general objective criteria must be applied to the product under discussion in order to become accepted as a proof (Heinze, 2010).

From this perspective, argumentation incorporates the construction of conjectures and proofs (Pedemonte & Buchbinder, 2011). Pedemonte (2007) termed the argumentation that contributes to the construction of a conjecture a "constructive argumentation" and the argumentation that justifies a conjecture a "structurant argumentation" (p. 390). Analogous to these terms, we use the notations of conjecturing respectively proving.

Different perspectives on the relationships

In the past, the relationships between argumentation, conjecturing and proving has been discussed from different points of view (Stylianides, 2007). Some researchers follow the work of Garuti, Boero, and Lemut (1998) by considering "the phenomenon of (possible) continuity between the production of a conjecture and the construction of its proof" (Boero, 1999; p. 5-6). This continuity is termed *cognitive unity*. It highlights that an argumentation in which a conjecture is produced can be extended to construct a proof by organizing the previously generated arguments into a deductive chain. It is based on the assumption that there is a close link between the nature of objects, the relations between objects, and the inferences used in both as well as the mental cognitive activities that arise during the conjecturing phase and the proving phase. Some researchers argue that proof is more "accessible" for students if some informal arguments such as drawings, examples, or theorems related to the proof have already been explored in the argumentation supporting the conjecture (cf. Pedemonte & Buchbinder, 2011).

Boero, Garuti, and Lemut (2007) observed that the dynamic conjecturing processes that led to the production of a conjecture can serve as thread, which has to be identified and then can be followed to build up a proof. Even though they pointed to the similarities between the processes of exploration performed during the conjecturing phase and during the proving phase, they remarked that the dynamic exploration differs in its function within these two phases: on the one hand it serves as support to the selection and the specification of the conjecture, on the other hand it reinforces the logical connection that has been made between the single arguments.

In contrast to this perspective, some researchers recommend to distinguish between the concepts of *mathematical argumentation* and *argumentation in mathematics* and consider argumentation as an "epistemological obstacle to the learning of mathematical proof" (Balacheff, 1999, p. 7), whereas others emphasize that there is a "structural gap" between argumentation and proof (e.g., Duval, 1995) as in argumentation inferences are related to the content while proofs usually have a deductive structure (i.e., claims are deduced from data by applying inference rules) (Pedemonte, 2008). However, Pedemonte (2007) claimed that argumentation and proof can have the same structure, but that this "structural continuity" does - depending on the mathematical domain in which the two processes of argumentation and proving are performed - not always favour proof construction.

The activity of conjecturing

Regarding the term conjecturing, Koedinger (1998) developed a cognitive model of conjecturing that defines the processes of discovering, recalling, and problem solving as the superordinate goals, and conjecture generation and argumentation for or against it as the two major sub-goals of conjecturing (see Figure 1). This model suggested that conjecture generation and argumentation (including inferences that are based on inductive or deductive arguments) are connected to each other via the investigation of examples and counterexamples. Furthermore, it indicated that a proof itself also served as a mean to discover new conjectures, even though the discovery of a conjecture usually constitutes the result of inductive reasoning strategies. Lin, F. L., Yang, Lee, Tabach, and Stylianides G. (2012) point out that conjecturing captures the observation from examples, the construction of new knowledge, the transformation of prior knowledge, and the reflection on the conjectured constructs as well as on the conjecturing processes itself.

Since there is no widely agreed definition of conjecturing, we conceptualize it as the activity that a student or mathematician engages in to find, explore, and formulate a conjecture, including the processes of generating (counter-) examples, searching for patterns, extending a set of examples into a general argument (informal induction), testing the conjecture's limitations, and presenting it to colleagues or teachers.

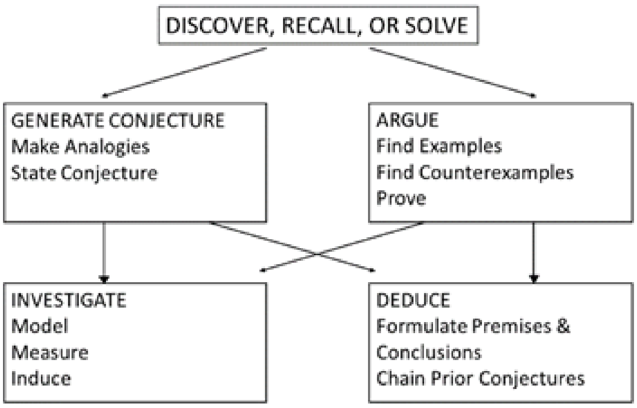

Figure 1: The goal structure for conjecturing and argumentation skills (Koedinger, 1998)

Toulmin's model as a tool to compare argumentation and proof

Toulmin's model constitutes the most common approach to represent the whole structure of an argumentation. We will briefly introduce it: In Toulmin's basic model, any argument consisted of three elements: *claim* (C): it is the assertion or an opinion of a speaker, *data* (D): these are the facts that justify the claim, *warrant* (W): it is the inference rule, which links the data to the claim and gives the data general support (cf. Toulmin, 1958). The model demonstrates that a speaker usually starts an argument by proposing a claim. In a next step, the speaker uses data to justify the claim and the warrant, as a general valid rule, to support the specific data that lead to the claim (see Figure 2).

Three further elements may be necessary to describe the whole structure of an argument depending on the situation (Toulmin, 1958): A *backing* to strengthen the warrant, a *qualifier* that shows the degree of confidence in the claim or in the conclusion, and a *rebuttal* that expresses exceptions or conditions for the validity of a claim.

In the past, several studies have used Toulmin's model to evaluate and compare students' argumentations and their proofs (cf. Pedemonte, 2007; 2008) or to investigate aspects of mathematical learning with regard to explanation, justification, and argumentation in mathematics classrooms (e.g., Krummheuer, 1995; Yackel, 2001). We agree that Toulmin's model is an appropriate method that can be used to analyse both the structure of an argumentation that has already been constructed as well as how argumentation structures are developed during proving processes and peer collaboration.

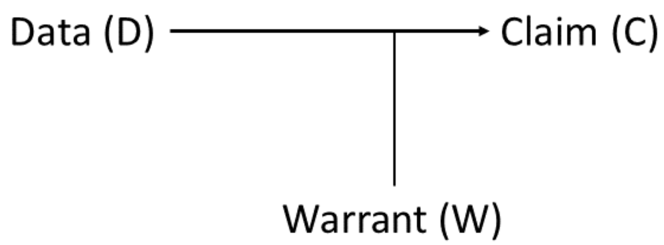

Figure 2: Toulmin's basic model. Visualizing the structure of an argument.

Summary

The presented debate demonstrated that different definitions of and perspectives on the relationship between proving, conjecturing, and argumentation "compete" with each other. Since Stylianides, G. J. et al. (2017) invite all researchers to be more explicit about the definitions one uses, we specify our perspectives on these three concepts as follows:

In this thesis, we consider conjecturing and proving as certain types of mathematical argumentation. Following Otten et al. (2017) and the definition of proof outlined by Stylianides, G. J. et al. (2017), a proof consists of a deductive chain of arguments using already stated definitions or propositions that determine a mathematical claim. In mathematics, proof is deductive, but the processes of attempting to construct a proof (what we call proving) as well as to discover and formulate a conjecture (what is termed conjecturing) are often characterized by informal, empirical argumentation. Conjecturing and proving processes incorporate several mental and physical actions (Selden, A. et al., 2010). Furthermore, we follow the conceptualization of argumentation proposed by Kollar et al. (2014) that is based on the assumption that argumentation, and thus also conjecturing and proving, comprise an individual-mathematical and a social-discursive component. From this perspective, the individual-mathematical component is related to phases of exploration and systematization (cf. Boero, 1999; Selden & Selden, 1995). The social-discursive component refers to the generation and exchange of well-warranted arguments (cf. Kollar, Fischer, & Slotta, 2007; Leitão, 2000) as well as to the social process of accepting these arguments as proof (cf. Heinze, 2010; Yackel & Cobb, 1996). Supposing that conjecturing and proving are interrelated and an important mechanism in constructing new knowledge, it seems to be an effective teaching strategy not to provide students with an initial conjecture, but to present open-ended problems that allow them to formulate their own conjectures, and then to prove those conjectures (Lin et al., 2012).

2.2.2 Situations in the context of mathematical argumentation

Mathematical argumentation activities and the outcomes of these activities occur in different (learning-) situations, which may explain why there exist various definitions of and perspectives on argumentation.

Situations in the context of mathematical argumentation include the social environment in which the argumentative discourse is embedded, the content area and the type of task that represent the argumentation problem, as well as the complexity of the task that determines which argumentative activities have to be employed to solve the task.

2.2.2.1 The social environment

The social environment comprises the mathematical community that often allocates a specific role to argumentation and proof, sets up learning goals in the context of mathematical argumentation, and establishes criteria and norms to which the acceptability of a proof is adhered.

There are distinct views on the role of argumentation and proof within mathematical learning and what makes a proof acceptable (Hanna, 2000). It can be assumed that a person behaves differently in the case of being requested to construct a proof to explain a mathematical statement than in the case of being requested to construct a proof to discover a statement. Moreover, the criteria and norms about what makes a proof acceptable depend on what educators and mathematicians expect from their students or peers. These influence mathematical practice (in particular, all activities that mathematicians engage in with regard to poof, including proof construction, proof reading, and proof presentation) and the teaching of proof in mathematics education. Yackel and Cobb (1996) introduced the notion of *sociomathematical norms*. These are normative aspects that emerged interactively and regulate mathematical argumentation. Sociomathematical norms determine “what counts as an acceptable mathematical explanation and justification” (p. 461) and are defined by the corresponding mathematical community. The taken-as-shared basis that has been established within this community includes these sociomathematical norms and serves as background that set which data and warrants legitimize one’s conclusions. Dawkins and Weber (2017) claimed that norms of proof mirror some of the features that mathematicians suspect to be necessary or desirable in the generation of new mathematical knowledge. In their paper, they considered the following four values: “(1) Mathematical knowledge is justified by a priori arguments. (2) Mathematical knowledge and justifications should be a-contextual and specifically be independent of time and author. (3) Mathematicians desire to increase their understanding of mathematics. (4) Mathematicians desire a set of consistent proof standards” (p. 128).

In particular, value (2) has to be regarded critically. On the one hand, there is a consensus that the values and norms that specify acceptability criteria depend on the mathematical community. On the other hand, there is a demand that the correctness of the proof can be evaluated independently from context and the person who created it. Mathematicians’ values and norms apprise and limit to some extent how proving practice should proceed.

However, proofs in the context of school mathematics are not, and cannot be, replications of proofs produced by expert mathematicians (e.g., Dawkins & Weber, 2017; Harel & Sowder, 1998; Weber, Inglis, & Mejia-Ramos, 2014). Within the framework of school mathematics, some statements are expediently and publicly used without further justification. What kind of statements can be used without further justification for the development of a proof, can vary between classroom communities (Stylianides, A. J., 2007).

In addition to the expectations of the mathematical community (representing an institution) and its established norms, against which arguments can be judged, it makes a huge difference whether students engage in social discursive argumentation practices or whether argumentation constitutes an individual activity. Social discursive argumentation practices require collaborative argument construction, including the critical discussion of ideas with others and the joint consideration of complex mathematical problems (Mueller et al., 2012). Students have to learn to collaborate effectively and to use the “exploratory talk” constructively (Mercer, Dawes, Wegerif, & Sams, 2004), since both can positively contribute to the development of mathematical arguments (e.g., Mueller et al., 2012). Presenting arguments to colleagues, evaluating colleagues' arguments, and trying to understand and learn from the shared arguments is a central part of argumentation practice (Dawkins & Weber, 2017).

2.2.2.2 The content area, type and complexity of a task

Based on the findings of previous studies that knowledge of concepts, facts and procedures connected to the specific content is a statistically significant predictor for students' performance in proof construction situations (e.g., Chinnappan, Ekanayake, & Brown, 2012; Sommerhoff, Ufer, & Kollar, 2016; Ufer et al., 2008), it can be said that the specific content area of the proof task plays a crucial role. In their study on proof schemes, Harel and Sowder (1998) started from the assumption that the nature of the task influenced what the students focused on and what processes they employed to gain certainty. The study of Mejía-Ramos and Inglis (2009) was also grounded on the hypothesis that variations in the task contexts could affect different behaviours. Their bibliographic study aimed to explore the different task-dependent argumentative activities that were associated with the notion of proof. Three types of activities could be distinguished with regard to the tasks that were frequently used in mathematical practices: the construction of novel arguments, the reading of given arguments, and the presentation of arguments. Each of these proof activities requires certain sub-processes depending on the given conditions and the intended goals. Mejía-Ramos and Inglis (2009) claimed that the comprehension of mathematical arguments (as a sub-process of argument reading) and the presentation of arguments, or at least parts of them, are the main activities involved in the assessment of undergraduate mathematics students' argumentation skills. However, these activities are often underrepresented within the literature. A complete categorization of argumentation processes with regard to the given conditions and intended gaols does not yet exist, but one could, for instance, include the conjecturing processes.

A further aspect that characterizes argumentation- or proof-situations is the complexity of a task. Ufer, Heinze, and Reiss (2009) showed that proof construction problems that require more than one step are usually more challenging for learners than one-step proofs. The complexity of a task regulates whether automatized reasoning strategies could be applied or whether an argument has to be constructed within the base of one's conceptual knowledge. For instance, multi-steps proofs require one's ability to recall, apply and connect different concepts and definitions, as well as planning and coordination processes.

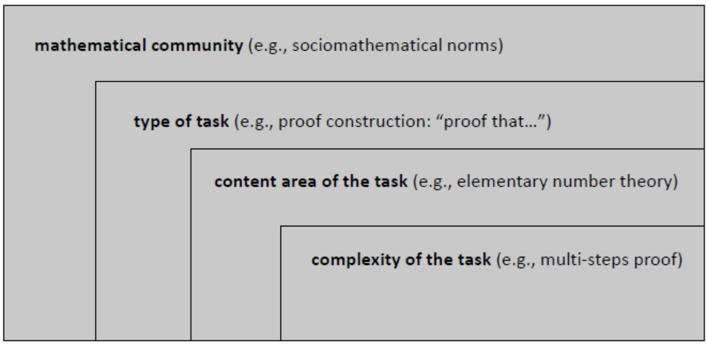

Figure 3: Situation-specific factors of mathematical argumentation and proof.

We conclude that argumentation has always to be considered in the context of the situation in which it is embedded for the following reasons: i) the success of argumentation is determined by the norms and criteria that the mathematical community specifies ii) the processes as well as the resources necessary for success depend on the content area, the type, and the complexity of a task. Figure 3 summarizes all the situation-specific factors.

2.2.3 Individual resources

Current discussions in educational research emphasize the importance of modeling human resources (including a person's stable and trainable skills and knowledge facets) that are responsible for mastering certain argumentation or proof situations. Researchers are interested in what knowledge facets are needed, and how various types of cognitive and affect-motivational resources direct one’s proving processes (Selden, A., Selden, & Benkhalti, 2018).

The results of several researchers have indicated that argumentation respectively proving is a knowledge-intensive activity (e.g., Chinnappan et al., 2012; Ufer et al., 2008). Ufer et al. (2008) empirically investigated the impact of declarative and procedural geometrical knowledge as well as mathematics-related problem-solving skills on students’ performance in proof construction situations. The findings of their study pointed out that more than 40% of the variance of students’ geometrical proof performance can be explained by these three cognitive predictors. Weber (2001) focussed on another knowledge facet of argumentation skills. He compared university students who had completed a course in abstract algebra with doctoral students doing research on this content area. It was observed that the doctoral students were able to make better use of strategic knowledge. The exploratory study of Weber and Alcock (2004) supported this result. They discovered that the doctoral students were able to choose a strategically better starting point to prove that two groups were isomorphic by examining the algebraic properties that preserved by isomorphism, whereas the undergraduate students immediately focused on the cardinality of these groups and thus were not able to construct a proof. Weber and Alcock (2004) interpreted this observation as an indicator for the strong impact of strategic knowledge on proving performance. The interview study of Heinze and Reiss (2003) has demonstrated that methodological knowledge about proof schemes, proof structures and the chain of conclusions, as a further cognitive resource, becomes particularly important when students are requested to evaluate correct and incorrect proofs. Sommerhoff (2017) investigated the impact of six underlying cognitive resources (conceptual and procedural mathematical knowledge, strategic knowledge, methodological knowledge, problem solving skills, metacognitive awareness, conditional reasoning skills) on students’ performance in proof validation and proof construction situations. Out of these six cognitive resources, conceptual knowledge as well as strategic knowledge showed a significant influence on the proof validation performance. Procedural and conceptual mathematical knowledge as well as strategic knowledge were predictive for the performance in proof construction situations. The strong impact of procedural knowledge on proof construction performance, but low relevance within proof validation made the major difference between both situations. Since prior findings regarding the effects of methodological knowledge and problem-solving skills could not be replicated in this study (cf. Chinnappan et al., 2012; Ufer et al., 2008), it can be assumed that different argumentative situations (in particular, different communities and content areas of tasks) require different cognitive resources.

Several researchers have conceptualized the activity of proof construction as a problem-solving task (e.g., Furinghetti & Morselli, 2009; Weber, 2005). From this perspective, it appears obvious to take metacognitive knowledge, beliefs, and self-regularity skills as further predictors for the successful outcome into account (cf. Corte, Verschaffel, & Op't Eynde, 2000). Selden, A. and Selden (2013b) also emphasized that beliefs about one's own ability to succeed in a specific situation as well as persistency (both affect-motivational resources) may play an important role on the success of university students' and mathematicians' proof construction processes.

Based upon these results, it can be expected that different cognitive resources (and affect-motivational resources) shape students' argumentation and proving processes as well as their performance in argumentation tasks. However, most researchers only have investigated the impact of specific resources on students' performance, disregarding the processes as a link between them.

2.2.4 Conceptualizing mathematical argumentation skills

Mathematics students usually demonstrate their skills and what they have learned during the semester by solving proving tasks (e.g., Selden, A. & Selden, 2008). Koedinger (1998) argued that the performances on conjecturing and proving tasks are the results of specific skills and knowledge facets. Therefore, students' written proof attempts are used as an important component in evaluating their skills and their conceptual understanding (Selden, A. et al., 2010).

From this perspective, mathematical argumentation tasks respectively proof problems determine the specific situational demands that can be mastered by individuals with a certain level of mathematical argumentation skills (cf. Koeppen et al., 2008).

Following several researchers, we conceptualize mathematical argumentation skills as "latent traits […] [that] cannot be directly observed but have to be inferred from observable behavior" (Blömeke et al., 2015, p. 3). Observable behaviour includes processes and performance in specific situations, which both can be judged against criteria to determine whether particular levels of skills have been reached (cf. Blömeke et al., 2015; Koeppen et al., 2008).

Prior findings in research, based on quantitative (e.g., Chinnappan et al., 2012; Sommerhoff et al., 2016; Ufer et al., 2008) and qualitative studies (e.g., Mejía-Ramos & Inglis, 2009; Selden, A. et al., 2010; Selden, A. & Selden, 2013a) have shown that the performance in argumentation tasks depends on several cognitive (e.g., Chinnappan et al, 2012; Sommerhoff et al, 2016; Ufer et al., 2008) and affective-motivational resources (e.g., Selden, A. et al., 2010; Selden, A. & Selden, 2013a), as well as on the situational demands (e.g., Mejía-Ramos & Inglis, 2009; Sommerhoff, 2017).

Furthermore, it can be assumed that the resources students bring to an argumentative situation guide the processes of generating arguments and proofs and that the performance is influenced by both processes and resources (cf. Carlson & Bloom, 2005).

Figure 4 presents our general model of mathematical argumentation skills: In response to the situational demands, underlying resources (learning prerequisites) are stimulated and used to implement the argumentation processes (consisting of a sequence of goals, mental and physical actions) that lead to the final product. The final product itself represents the performance in the specific situation.

The argumentation skills (including cognitive and affective-motivational resources) and the performance constitute a linked system, cobbled together by the argumentation processes that may work as mediators.

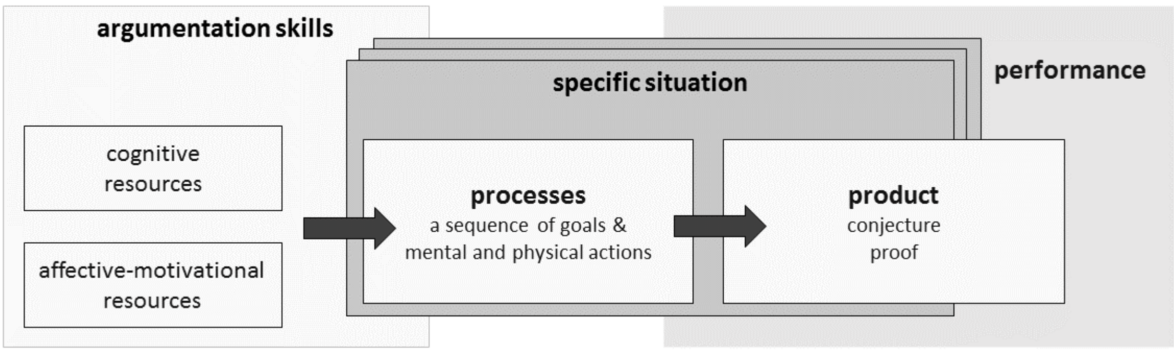

Figure 4: Conceptualization of mathematical argumentation skills.

In the past, several researchers have developed frameworks that describe the argumentation processes and thus demonstrate how mathematicians execute throughout the proof construction process (e.g., Boero, 1999; Schwarz, Hershkowitz, & Prusak, 2010). In the following, we will present some of these frameworks.

2.2.5 Argumentation processes

This section provides an overview about the frameworks and findings of previous research on (mathematical) argumentation processes. As the focus in this dissertation is mainly on conjecturing and proving (as specific types of argumentation), we will primarily discuss the frameworks of proof construction that include phases of exploration and systematization.

2.2.5.1 The framework of scientific reasoning and argumentation

We start with a framework that is not specific to mathematics, but that covers current research on and models of scientific reasoning and argumentation (SRA) from various scientific disciplines. It is the framework proposed by Fischer et al. (2014) who suggested to differentiate eight processes (that are called *epistemic activities*) to characterize reasoning and argumentation in any kind of scientific discipline. These processes are termed: *problem identification, questioning, hypothesis generation, construction and redesign of artifacts, evidence generation, evidence evaluation, drawing conclusions* and *communicating and scrutinizing.*

Out of these eight epistemic activities, one can argue that hypothesis generation, evidence generation, evidence evaluation, drawing conclusions, and communicating and scrutinizing are the core of argumentation and proof construction processes in the context of secondary school and university mathematics. The framework describes these activities in the following way:

- Hypothesis generation is the process of formulating a hypothesis (sometimes as a possible answer to the question under study) with regard to obvious models, available theoretical tools, or empirical evidence.
- Evidence generation comprises empirical and formal approaches to gather evidence. However, within the domain of mathematics, empirical evidence generated by, examining examples, constitutes only preliminary evidence. A deductive chain of arguments has to be constructed to establish the validity of a statement based on the definitions and axioms of a mathematical theory.
- Evidence evaluation is the process of assessing the degree to which an argument as piece of evidence supports a particular claim by taking certain norms (e.g., sociomathematical norms) into account.
- Organizing and integrating different pieces of evidence as well as re-evaluating the initial claim by critically analysing data and warrants are summarized as the activity of drawing conclusions.
- Communicating and scrutinizing describes the process of sharing and presenting one's results. This last activity is strongly related to the social character of argumentation.

Mathematics researchers may also be confronted with the other three epistemic activities: For instance, they may discover a discrepancy or shortcoming regarding the available explanation of a specific mathematical problem (problem identification), they may formulate one or more initial questions as a driving force for their research (questioning), and they may create a prototypical object or an axiomatic system that introduces a new mathematical structure (construction and redesign of artifacts). It appears obvious that students are less faced with these types of activities as they usually receive a concrete problem, a well-defined question and are requested to apply and combine definitions and statements based on the axiomatic system.

We conclude that the framework by Fischer et al. (2014) that is based on numerous theoretical considerations provides a detailed picture of the activities (phases) that can occur in mathematical argumentation and proof construction processes in different contexts, but does not allow the derivation of hypotheses on how these activities have to be employed to achieve high performance.

2.2.5.2 The four phases of problem-solving

Conceptualizing proving as problem solving (e.g., Weber, 2005) leads to the four phases already outlined in Polya (1945): *understanding the problem, devising a plan, carrying out the plan,* and *looking back.*

- *Understanding the problem*: implies to understand all the words used in stating the problem, to recognise what one is asked to find or to show, to draw figures, and to separate the various parts of the condition.
- *Devising a plan:* incorporates the finding of a connection between the givens and the unknown as well as the choice of an appropriate strategy to solve the problem.
- *Carrying out the plan:* covers the processes of persisting with the plan that one has chosen and discarding and choosing another plan if it continues not to work.
- *Looking back:* involves checking the results and thinking about whether the results could have been derived otherwise.

This framework, which suggests strategies for attacking problems in mathematics classes, has already been used in a slightly amended form to describe the processes involved in proof construction (cf. Furinghetti & Morselli, 2009; Selden, J. & Selden, 1995; Selden, A. & Selden, 2009; VanSpronsen, 2008)

2.2.5.3 Frameworks of proof construction

Research that refers to the (possible) continuity between the production of a conjecture and the construction of its proof is usually conceptually linked to the phase model proposed by Boero (1999) that deals with the role of argumentation in the domain of mathematics and that distinguishes between phases of exploration and systematization.

2.2.5.4 Boeros' expert model

This expert model consisted of six phases that are not intended to be interpreted as a linear sequence.

- The first phase involves the exploration of the problem situation and the identification of patterns. Processes associated with this phase are performed with the goal of generating a conjecture.
- The formulation of the statement to be proved constitutes the second phase.
- Exploring the content by questioning the limits of the conjecture and by applying semantic (or sometimes even syntactic) reasoning strategies represents the third phase.
- The fourth phase is characterized by selecting appropriate arguments that serve as supporting warrants for the conjecture as well as enchaining these arguments into a coherent deductive line.
- Organizing the enchained arguments into a proof that meets the corresponding mathematical community's expectations and sociomathematical norms forms the fifth phase.
- The process of achieving a completely formal proof is covered by the sixth phase which, in most cases, is skipped either due to its irrelevance or impossibility within the context of school or university mathematics.

Boero (1999) clarified that the occurrence of each phase depends on the context in which the argumentation is embedded (including the culture, community and type of task), and distinguished between the *private side of mathematicians' work* (related to phase I and III – phases of exploration) and the *public side* (especially represented by phase II, V, and VI – phases of systematization). Moreover, he emphasized the importance of differentiating between conjecturing and proving as processes on the one hand and theorems respectively proofs as the resulting products on the other hand. Boero (1999) refers to the process-product character of proof by claiming that a proof as the final product of one's mathematical argumentation processes has to meet certain formal criteria in order to be accepted as a proof, but the conjecturing and proving processes used to generate this proof do not. Therefore, when proving a conjecture, one can first generate an informal argument trying to convince oneself about the validity of a conjecture and then use this informal argument as a substructure to produce a proof (e.g., Garuti et al., 1998; Weber & Alcock, 2004).

Heinze and Reiss (2007) extended Boero's model by adding the phase of *acceptance by the mathematical community* to take the social act of proving (cf. Manin, 1977) into account.

2.2.5.5 Proving as a problem-solving pathway

Furinghetti and Morselli (2009) adapted Polya's problem solving framework to the process of proof construction by regarding the proving process as a pathway including processes of exploration and systematization and consisting of these four phases:

- They described the *first phase* as becoming acquainted with the given task by carefully reading and reformulating the text that represents the proving problem and by attempting to reach a logical understanding. Drawing pictures and generating examples may be part of the reformulating processes that bridge to the phase of developing a plan.

- The *second phase* involves the thoughtful choice of proving strategies, methods, and representations. They emphasized the importance of finding a predictable representation by switching from one representation to another and of going back to previous steps to overcome deadlocks. Furinghetti and Morselli pointed to Weber and Alcock (2004) by claiming that an arithmetic/ algebraic representation may be such a predictable representation and thus the starting point for the syntactic proving strategy.

- The monitoring and use of proving strategies was regarded as the *third phase*: syntactic proving strategies incorporate the application of definitions and procedures such as manipulating symbols in an automatic-like style; semantic proving strategies are based on meaningful instantiations to guide the formal inferences and require the transition from informal to formal arguments (cf. Weber & Alcock, 2004).

- The *fourth phase* of the proving pathway was characterized as evaluation and reflection phase including processes such as revising the employed proving strategy, checking the chain of deductions inferred during the third phase, as well as refining the language by the use of formal mathematical symbols.

This theoretical framework points to the relevance of self-regulatory activities (planning, monitoring, and reflection) within proof construction. It was used to guide the analysis of the written proof attempts of two unsuccessful mathematics students and helped to identify the difficulties that students' encountered in proving a statement.

2.2.5.6 Processes related to the formal-rhetorical and the problem-centered part of a proof

Selden and Selden (1995) suggested to differentiate between the processes related to the *formal-rhetorical* part and the processes related to the *problem-centered* part of a proof.

- The formal-rhetorical part of a proof consists of "unpacking and using the logical structure of the statement of a theorem and associated definitions" (Selden, A. et al., 2010, p. 200). This includes starting the proof by writing the premise at the beginning, leaving space for the main body, and writing the conclusion at the end of the proof. The following steps are unpacking the conclusion, selecting the relevant definitions, and adapting the symbols used in the definitions to the theorem that has to be proven (Selden, A. & Selden, 2011). Constructing a hierarchical structure of a proof can be considered as calling a schema, and the processes involved are often executed automatically (Selden, A. & Selden, 2009).

- Filling the space for the main body of the proof refers to the problem-centered part, where some "exploration" and "brainstorming" processes gain in importance. They

> claimed that problem-solving strategies such as drawing diagrams, reflecting on the results of prior activities, or trying to remember an example are employed to link different concepts and to develop an idea for how to proceed (Selden, A. et al., 2010; Selden, A. & Selden, 2011), both required to generate the problem-solving part of a proof. Furthermore, they observed that there are propositions for which constructing the hierarchical structure of a proof can be very useful in uncovering the "real problem" to be solved in the remaining proof (Selden, A. & Selden, 2009).

In their teaching of proof, Selden, A. and Selden (2008) addressed this distinction of processes related to the formal-rhetorical part and those related to the problem-solving part to overcome students' difficulties with proof construction and to enhance their proving performance.

The frameworks of proof construction related to problem solving suggest that much can be gained by thinking about what may work and by reflecting on what has worked respectively not worked. Furthermore, they point out that processes of systematization and exploration have to be used to develop the structure and ideas that build a formal proof and thus, are conceptually linked to Boero's expert model.

2.2.5.7 The cognitive model of conjecturing

The cognitive model of conjecturing proposed by Koedinger (1998) consists of four components: *generate conjecture, investigate, argue,* and *deduce* (see Figure 2). Conjectures may be generated by the investigation of examples, by drawing analogies to familiar problems, or even by deduction. These components may also be relevant for testing a conjecture and generating evidence for or against it. Conjecture generation comprises identifying patterns and commonalities between examples, ensuring that the conjecture is consistent with the empirical evidence that has been created before, and affirming that it goes beyond simply replicating the premise. Investigation involves exploring examples or, in content areas such as geometry, constructing models, measuring objects such as segment lengths, angles, and areas, and inducing any relationships that appear to be invariant. Argumentation describes the process of generating empirical or deductive arguments. Deduction, as used in Koedinger's model, refers to the formulation of conclusions and to the application of theorems in order to generate a deductive chain from the givens to the conclusions.

Koedinger's model was developed to describe the observations of middle school students' performance on a conjecturing task. According to the school context and geometric content area, it primarily focusses on exploratory processes, but also includes formal-deductive processes.

2.2.5.8 Summary

From the comparison of the frameworks presented above, we inferred that some of these frameworks have been developed to enhance the teaching and learning of proof (the framework of Selden, J. and Selden (1995)) respectively problem solving (Polya, 1945). Other frameworks have been created to describe school students (the framework by Koedinger, 1998), university students (the framework by Furinghetti and Morselli (2009)), and expert mathematicians (the framework by Boero (1999)) proof construction processes or argumentation activities (the framework by Fischer et al. (2014)).

The frameworks of proof construction have in common that they emphasize the crucial role of dynamic exploration (cf. Boero et al, 1996) and systematization: Furinghetti and Morselli's work (2009) by comprising the phases of understanding the problem and developing a plan as phases of exploration and carrying out the plan and looking back as phases of systematization; Selden, J. and Selden (1995) by referring to the problem-centered part and the formal-rhetorical part of a proof; Koedinger (1998) by introducing the components of conjecture generation, investigation, and argumentation as exploratory activities and deduction as a systematization activity; and Boero's model (1999) by containing phases related to the private side of a mathematician (phase I and III) and the public side (phase II, V, and VI) .

During the epistemic activities of problem identification and evidence generation, explorative processes such as studying and testing examples may occur as well (Fischer et al., 2014). The dichotomy of proof suggested by Selden and Selden (1995), Koedinger's model (1998), as well as Borero's model (1999) include phases of systematization, but they do not explicitly bring up the processes of communicating and scrutinizing (cf. Fischer et al., 2014) respectively of looking back (Polya, 1945).

These frameworks can be used to describe argumentation, conjecturing, and proving processes. They demonstrate that argumentation processes are a sequence of sub-goals (phases) including physical and mental actions. We conclude that phases of exploration and systematization are required to communicate arguments precisely. Yet, what does a *good* exploration phase or systemization phase look like? Regarding this question, we observed some hints in the literature. For instance, we have inferred from the work of Koedinger (1998) that testing a conjecture with multiple examples, searching for a counterexample, as well as checking the inferences can be considered as characteristics that describe how conjecturing processes should be. The framework of Furinghetti and Morselli (2009) also pointed to individual quality features that characterize the proof construction process such as carefully choosing appropriate representations or going back to previous steps in the case of impasses.

However, an overview of which *process characteristics* are actually relevant for the production of an *interesting conjecture* (that goes beyond repeating the premise) and *a correct and normatively acceptable proof* does not yet exist. Even though the success of argumentation processes is determined by the quality of the final product, it is still unclear how "good" argumentation processes can be described. We see the need (i) to summarize common claims and findings of previous research on *good* argumentation processes and (ii) to empirically investigate which of the theory-based process characteristics are predictive for the quality of the resulting product.

The frameworks outlined above mainly address the argumentation processes of a person working individually. As argumentation is often embedded in social contexts (Balacheff, 1998; Yackel & Cobb, 1996), one may argue that the social-discursive argumentation processes of a person working in dyads or groups should also be taken into account when attempting to find a set of process characteristics that describe *good* (collaborative) conjecturing and proving processes.

2.2.6 Argumentation as a social practice – the role of collaboration

Some argumentation processes such as communicating and scrutinizing arguments are social-discursive in nature, others such as generating hypothesis and evidence generation may benefit from collaborative argumentation (cf. Fischer et al., 2014). When undergraduate students learn to construct proofs, the social perspective on argumentation, including the social nature of argumentation as well as the role of interaction and transactive reasoning, should be considered as well (Blanton & Stylianou, 2014). In the first part of this section, we will describe the social nature of argumentation and proof, followed by a section on the mechanisms of collaboration that may (positively) influence the generation of mathematical arguments. To approach this issue, we will expand our literature search to studies related to research in educational psychology and the Learning Sciences (esp. Computer-Supported-Collaborative-Learning research).

2.2.6.1 The social nature of argumentation

Besides their other functions, mathematical argumentation and proofs are means of communication (e.g., Hanna, 1990; Villiers, 1999). Students or mathematicians engage in proving to generate new knowledge and to justify or explain to their peers why a statement is true. It is the teacher or the peer group that judges whether an argument is a proof (e.g., Manin, 1977; Stylianides, G. J. et al., 2017). Within this perspective, the focus tends to be on the social processes that play a particular role in the acceptance of new results by the mathematical community. Consequently, educators have started to pay more attention on the concept of proof as a “convincing argument” (Hanna, 2000). It cannot be said that the processes of ascertaining and persuading (Harel & Sowder, 1998), in the sense of removing one’s own and others’ doubts (cf. Mariotti, 2006), as well as the social processes of evaluating the arguments of others and checking their logical integrity make mathematics less objective or true; rather, the modern view of the logical truth or validity of a mathematical statement relative to a reference theory has to be taken (Ernest, 1998) and proofs have to be considered as arguments that meet the norms shared by the respective community. This means a proof is always linked to the social context in which the proof occurs. It is a specific type of discourse, a form of interaction that is based on shared meanings (cf. Villiers, 1999). Therefore, teachers and instructors should give their students the opportunity to engage in the activity of proving as it is practiced in the mathematical community, including using proof to raise debates about the truth of a conjecture and to negotiate the meanings of concepts as well as (implicitly) the criteria for an acceptable proof (e.g., Alibert & Thomas, 1991; Villiers, 1999). Encouraging students to participate in mathematical discussions and to solve proof-related tasks collaboratively, leads to the question of how collaboration can be effectively used as a resource for generating proofs, and which characteristics make collaboration finally effective.

2.2.6.2 Mechanisms of collaboration

The phenomena of collaborative learning and problem solving and how they are influenced by one's cognitive and affect-motivational resources or by the use of collaboration scaffolds has been put forward in psychology and the Learning Sciences research (esp. Computer-Supported-Collaborative-Learning research) (e.g., Kopp & Mandl, 2011; Schwaighofer et al., 2017; Stahl, 2010; Vogel et al., 2016; Webb, 1982; Weinberger & Fischer, 2006). There is some evidence that collaborative learning or problem solving is not necessarily effective (especially in unstructured leaning situations) (e.g., Andriessen & Schwarz, 2009; Gillies, 2004) or at least, an advantage over individual learning (e.g., Barron, 2003; Yetter et al., 2006). Some experimental studies that directly compared the work of individuals and groups have shown that groups often outperform the average individual, but not when the level of pooled outcomes of "competent" individuals working alone were also taken into account (e.g., Schwartz, 1995). One reason might be that students often have problems engaging in deep-elaborative discourses when working together (e.g., Kollar et al., 2007; Vogel et al., 2017).

Research has demonstrated that particularly argumentative dialogues are essential in collaborative learning. Students that were instructed to engage in argumentative dialogues during collaboration were found to reflect better conceptual understanding in evolutionary theory than those who have not been instructed (Asterhan & Schwarz, 2007). Chi and Wylie (2014) argued that some studies were not able to provide evidence for the advantages of collaborative learning or problem solving on the reason that these studies did not distinguish between *individual dialogue* and *joint dialogue* respectively *interactive dialogue* patterns. In their paper, they emphasized that individuals can mainly benefit from dialogues when these dialogues are truly *interactive*. Interactive (also called *transactive* (Teasley, 1997)) dialogues involve the mutual exchange of ideas between the participants and lead to new ideas that go beyond the ideas one would be able to generate alone. Within interactive dialogues, the participants make substantive content-related contributions, such as generating arguments to support a position, asking critical questions, and elaborating on each other's comments. Webb (1989) pointed out that the level of elaborated explanations is likely to be positively related to one's achievement. They claimed that providing highly elaborated explanations requires the activities of clarifying and reorganizing the learning material that, in turn, may improve one's understanding. These assumptions are consistent with the model of Wecker and Fischer (2014) indicating that the cognitive processes influence the social activities (see Figure 5). However, it remains an empirical question to what extent individual-cognitive and social-discursive processes are related to each other.

Regarding mathematical argumentation, Mueller et al. (2012) suggested a framework for analysing collaborative mathematical argument construction by differentiating three types of collaboration, namely co-construction, integration, and modification. Co-construction of arguments implies that students collaborate in a back and forth manner by negotiating various positions until a mathematical argument is jointly built. Integration occurs when the argument of a learner is strengthened by assimilating the ideas and arguments produced by his or her learning partners. The third type of collaboration is that of modification incorporating the processes of challenging and evaluating the arguments of others. Results indicated that all three modes of collaboration influence students' building of mathematical arguments. The study of Goos, Galbraith, and Renshaw (2002) showed that transactive reasoning in small group peer discussions affected students' metacognitive processes that are crucial for mathematical problem solving. Similar to Goos et al. (2002), Blanton and Stylianou (2014) found that tansactive reasoning in mathematics classroom discourses as a habit of interaction encourage students' proof construction processes and their proof understanding.

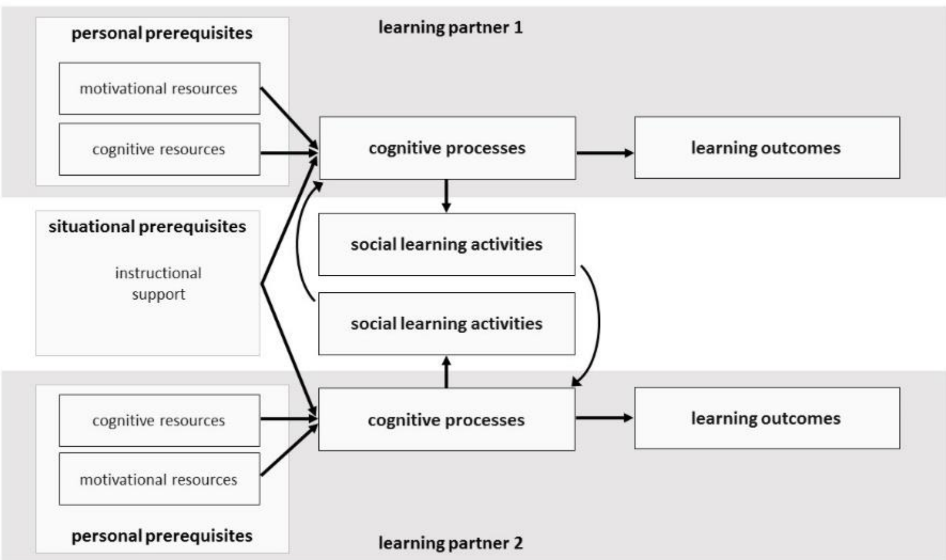

Figure 5: Interplay between personal and situational prerequisites, cognitive processes, and social learning activities within collaborative learning (Wecker & Fischer, 2014).

To identify the mechanisms and processes that make collaborative group settings effective, the model proposed by Wecker and Fischer (2014) may provide a theoretical basis for such analyses. This model aims to clarify how the learning outcomes of groups arise by focussing on the interactions between personal and situational prerequisites, cognitive processes, and social learning activities. Within this model, learners are considered as individuals that engage in their own, private cognitive processes. These cognitive processes underlie motivational and cognitive resources, and are influenced by situational context factors. One's cognitive processes have an impact on one's learning outcomes as well as on one's social activities that contribute to the collaborative discourse. In addition, the model suggests that the social learning activities of an individual affect the dyadic partner's cognitive processes (see Figure 5).

We conclude that (mathematical) argumentation can be viewed as social practice rather than a purely individual activity (cf. Fischer et al., 2014), especially from a perspective that conceptualizes proving as means for generating and communicating mathematical knowledge (e.g., de Villiers, 1990), and for establishing social norms with respect to proof (cf. Yackel & Cobb, 1996). Findings of prior research have shown that argumentation processes can be influenced by social interactions (e.g., Mueller et al., 2014; Wecker & Fischer, 2014). Collaborative mathematical argumentation requires developing a shared understanding of the proving problem and creating a common ground that provides the basis for the students' (and the teacher's) collaborative work (Staples, 2007). Exchanging arguments mutually and integrating the ideas of others appeared to be an important aspect of fruitful collaboration (e.g., Chi & Wylie, 2014; Mueller et al, 2014). From this point of view, it can be assumed that mathematical argumentation processes may benefit from fruitful collaboration.

Summary

In this thesis, we restrict ourselves to (collaborative) conjecturing and proving processes as specific types of argumentation. These processes are employed with the aim to generate hypotheses and to construct subsequent proofs. Other proof-related activities such as proof reading or proof presentation (cf. Mejía-Ramos & Inglis, 2009) were not investigated. We propose a model of mathematical argumentation skills that differentiates between learning prerequisites, argumentation processes, and the final product. Following the framework proposed by Blömeke et al. (2015), we assume that argumentation skills can be inferred from the argumentation processes and the final product representing the argumentation performance in a specific situation. Based on the model of Wecker and Fischer (2014) as well as on the theoretical considerations in Stahl (2010), we suppose that students working in collaborative dyads are active as individuals, as group participants, and as members of a broader community (e.g., the mathematics university community). From this perspective, students who are active as individuals are expected to employ their own individual-mathematical processes such as generating examples, formulating conjectures, applying definitions, selecting appropriate arguments, and organizing these arguments into a deductive chain. Students as group participants may share many of these processes with their learning partners. Therefore, social-discursive processes such as explaining one's own ideas and evaluating the arguments of others seem to be crucial as well. The mathematical community has created and accepted some criteria that are important for assessing the products of conjecturing and proving processes. The quality of these products determine the success of argumentation processes. We assume that the final products (i.e. the generated conjecture and proof) are directly affected by one's learning perquisites (that comprise several cognitive and affect-motivational resources) and that individual-mathematical and social-discursive processes work as mediating elements between the learning prerequisites on the one hand and the final products on the other hand.

3 The motivations behind defining process characteristics of collaborative conjecturing and proving

This thesis is mostly about the processes of conjecturing and proving and also takes the final product (i.e. the formulated conjecture and generated proof) into account. We focus on the process-product correlation with the aim to identify and define process characteristics of collaborative conjecturing and proving that are relevant for the success. In this chapter the following questions should be answered: Why would researchers be interested in process characteristics of conjecturing and proving? Which added value for research and practice can be expected? We will briefly describe several motivational aspects for defining process characteristics of collaborative conjecturing and proving form different perspectives.

Based on the assumptions of Meier et al. (2007) who developed a framework for assessing the quality of computer-supported collaboration processes, we suppose that researchers with interest in studying collaborative conjecturing and proving processes strive to answer three basic questions: 1) which characteristics of collaborative conjecturing and proving processes are relevant for success and 2) should therefore be observed? And 3) to what extend can different directly observable process characteristics be related to learning prerequisites?

3.1 Expanding theoretical knowledge about conjecturing and proving processes

The first aspect refers to expanding the theoretical knowledge about conjecturing and proving processes: In the past, many studies of argumentation and proof have examined errors and misconceptions in students' written proof attempts. Overall, students' proving performance at the secondary and tertiary level is primarily found to be weak. Selden, A. and Selden (2008) claimed that we need to know more than that a student can, or cannot, prove a specific theorem in a certain content area by, for instance, induction, deduction or contradiction. Mejía-Ramos and Inglis (2009) critically notated that the knowledge about students' proof-related activities is rare. It is still an open question what the relationships between proofs (as products) and the processes that mathematicians use to construct these products are (Douek, 2007). In general, there seems to be a need for a comprehensive view on conjecturing and proving in an effort to understand students' difficulties and the sources of these difficulties on the one hand, as well as the processes that positively influence students' conjecturing and proving performance on the other hand (cf. Harel & Sowder, 2007).

The performance in form of the resulting product can be evaluated relatively clearly according to certain criteria (cf. Miller, Infante, & Weber, 2018). This gives rise to the question of which process characteristics are associated with high-quality proof products and thus describe *good* (collaborative) conjecturing and proving processes. Furthermore, prior research has shown that the quality of the proof product strongly depends on learning prerequisites (e.g., Ufer et al., 2008). From this perspective, it is important to find out to what extent the process characteristics of conjecturing and proving mediate the relationship between learning prerequisites and the quality of the final product.

We refer to these process characteristics as individual-mathematical process characteristics of (collaborative) conjecturing and proving. Theoretical knowledge about individual-mathematical process characteristics could provide the basis for developing a model of *good* conjecturing and proving processes that researchers may use for further (empirical) investigations. As already discussed in previous chapters, conjecturing and proving are often regarded as social activities. Therefore, to identify which peer collaboration characteristics may be crucial for proof construction when proving is situated in a social context may open up new venues for research on the mechanism of learning mathematical argumentation and proof in and from peer collaboration (cf. Asterhan & Schwarz, 2009). We call the peer collaboration characteristics related to proof construction as social-discursive process characteristics of collaborative conjecturing and proving. To sum up, inferring process characteristics from the literature and defining those that are assumed to be observable and relevant for the success might be a starting point for following empirical analyses.

3.2 Using process characteristics to measure argumentation skills

The second aspect comprises the approach of using process characteristics to gain insight into students' or mathematicians' underlying argumentation skills. Since reaching a high reliability of assessing skills on performance-related tasks usually requires a huge number of items (e.g., Blömeke et al., 2015; Koeppen et al., 2008), it might be interesting to find out which characteristics of conjecturing and proving processes (beyond the final product) provide additional information about students' underlying argumentation skills. Argumentations skills as latent constructs cannot be directly observed but may be inferred from observable process characteristics of collaborative conjecturing and proving (cf. Blömeke et al., 2015). This means that process data could be examined to receive a more complete analysis of student's argumentation skills or to find out how students develop argumentation skills. A resulting question is how these process characteristics should be assessed (employing what kind of instrument) and how they can be operationalized.

In the literature, two types of measurements are distinguished: Processes can be measured and assessed at the same time they are occurring or as verbal or written representations of an activity taking place at an earlier date (cf. Shernoff & Kratochwill, 2003). It would be very instructive to have research on how school students, (advanced) university mathematics students, or expert mathematicians actually generate conjectures and proofs in real time to understand the temporal sequence of or the interplay between different processes. Even though such kind of studies already exist (e.g., Savic, 2015b), the research on conjecturing and proving processes based on real-time observations are comparatively rare (Selden, A. & Selden, 2013b). Most results about argumentation skills are inferred from written solutions. For instance, the research group around Reiss (e.g., Heinze, Cheng, Ufer, Lin, & Reiss, 2008b; Heinze, Reiss, & Franziska, 2005) studied the geometry-proof skills and conceptions of high school students by conducting multiple longitudinal surveys. The empirical results of these studies were derived from students' written responses to geometric test items. Students' written proof products were also used to identify mistakes and misconceptions (Selden, J., Benkhalti, & Selden, 2014) or to infer actions that might be beneficial for the construction of proofs (e.g., Selden, A. et al., 2010; Selden, A. et al., 2018).

We do not want to claim that analysing written responses to test items is detrimental because different types of proof (that are required to answer these items) may correspond to different skills and processes needed to create them (Selden, A. & Selden, 2008). Furthermore, the quality of the written proofs as results of the processes determines the success and thus the quality of these processes. Therefore, we assume that taking processes and the final product into account may be a promising approach for assessing mathematical argumentation skills. We propose an analytic model to measure argumentation skills (see Figure 6): argumentation processes that are enacted within a specific situational context become observable in the form of diverse process characteristics, learning prerequisites are represented by an individual's prior knowledge on proof, and the final product represents the individual's argumentation performance. The methods to measure each component may include real-time observations of the processes (assessed along diverse process characteristics), written proofs as representations of the final product (evaluated against diverse product criteria), and proof items presented in form of a paper-pencil test to capture prior knowledge on proof (respectively (learning-) prerequisites).

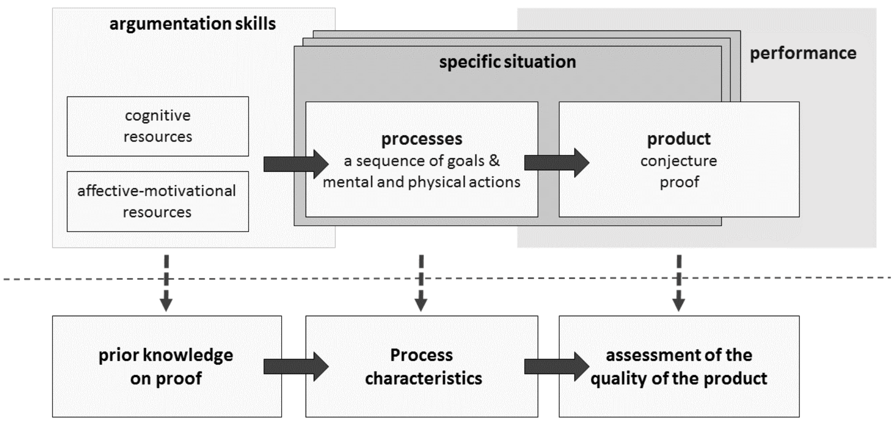

Figure 6: An analytic framework for measuring argumentation skills.

3.3 Using process characteristics as a diagnostic tool

The third aspect points to using process characteristics as a diagnostic tool in order to detect students' main deficiencies concerning the formulation of conjectures and the construction of proofs. Knowing which process characteristics are crucial for the success may help to identify where support is actually needed (Meier, Spada, & Rummel, 2007). Adaptive interventions and scaffolds to encourage students' (collaborative) conjecturing and proving skills may be developed on the basis of this knowledge. It allows to design tasks that may optimize the learning results by purposefully promoting and scaffolding those process characteristics that deemed to be important (cf. Schwartz, 1995). Consequently, students' may build awareness of the process characteristics that are considered to be responsible for successfully formulating conjectures and constructing proofs. This perspective conceptualizes process characteristics of collaborative conjecturing and proving as potential predictors of the quality of the resulting product (the quality of the formulated conjecture and generated proof). The additional focus on social-discursive process characteristics may even make it possible to stimulate fruitful collaboration patterns, which in turn may enhance the individual-mathematical process characteristics (cf. Mueller et al., 2015).

4 Aims and research questions of the three studies

As there is a general consensus on the importance of proofs in university mathematics, and in school mathematics (e.g., Harel & Sowder, 2007; Stylianides, A. J., 2007; Stylianides, G. J. et al., 2017), conjecturing and proving represent a major line in mathematics educational research (Sommerhoff, Ufer, & Kollar, 2015). Although the amount of research appears to provide a comprehensive view on conjecturing and proving, available empirical results about the activities involved that characterize *good* collaborative conjecturing and proving processes are rather weak. In the past, most studies about conjecturing and proving processes were based on small sample sizes using qualitative methods (e.g., Ellis et al., 2017; Savic, 2015b; Zazkis et al., 2016, 2015). Models of proof construction were inferred from introspective (e.g., Polya, 1945) or observational methods (e.g., Boero, 1999; Schwarz et al., 2010; Selden, J. & Selden, 1995)and were mainly premised on theoretical assumptions. In addition, most of these studies or models do not explicitly refer to the process-product character of proof (with the exception of Boero's model (1999)) or to collaboration. Until now, research on what constitutes "good" collaborative conjecturing and proving processes has not been specific enough to describe students' interactions during poof construction in terms of process characteristic that are relevant for the successful outcome (i.e. an interesting conjecture and a correct and normatively acceptable proof).

Building on both the conceptualization of mathematical argumentation skills as consisting of an individual-mathematical and social-discursive component as well as the process-product character of proof, the primary goal of this dissertation is to develop an instrument to describe and analyse collaborative conjecturing and proving processes from a mathematics educational and more domain-general social-discursive perspective on argumentation. We will use this instrument to study the "black box" of collaborative conjecturing and proving by addressing the two questions: "What is happening during the process of proof construction?" and "How do learning prerequisites affect individual-mathematical and social-discursive processes that may lead to a successful outcome?"

Yet, before developing such an instrument, it is important to know (1) which individual-mathematical and social-discursive activities are considered as process characteristics that constitute *good* (collaborative) conjecturing and proving processes from a theoretical point of view based on current literature. Study I (Chapter 5) is a literature review that approaches this issue and summarizes common claims and findings about conjecturing and proving processes. More specifically, this literature review aims to answer the following questions: (1.1) Which theoretical perspectives in mathematics-related educational research conceptualize characteristics of good conjecturing and proving processes? Do researchers discuss good conjecturing and proving processes from the proving as problem-solving, proving as convincing, or proving as a socially-embedded activity perspective? (1.2) Which conjecturing and proving processes are considered to be crucial for the production of interesting conjectures and normatively acceptable proofs? (1.3) Which of these proving processes are reported as more general sub-goals within conjecturing and proving? And which process characteristics are assumed to be helpful to achieve these goals? The results of these research question serve as theoretical foundation for defining a set of process characteristics of (collaborative) conjecturing and proving.

Further goals of this dissertation are to find out (2.1) how can such process characteristics be operationalized in the context of collaborative proving and conjecturing processes? Is it feasible to measure individual-mathematical and social-discursive process characteristics of collaborative conjecturing and proving reliably? This research issue requires the development of an instrument (including the operationalization of process characteristics inferred from mathematics educational and psychology research). This instrument could be used (2.2) to analyse the empirical structure of individual-mathematical and social-discursive process characteristics and to find out whether coherent dimensions of process quality can be identified, (2.3) to study to what extent both learners contribute equally to the performance of these process characteristics, and (3) to examine the mutual relations between prerequisites, process characteristics, and proof performance.

To investigate these broad range of research issues, we conduct an empirical study situated at the transition phase from secondary to tertiary education. This empirical study is broken down in two sub-analyses (Study II and Study III). Furthermore, this dissertation provides a “technical report” (Chapter 6) that presents the development of a high-inference coding instrument, a description of the rating scales and of the coding procedure, as well as a short explanation why we used this type of methodology.

Study II (Chapter 7) focuses on aspect (2.1) by systematizing and operationalizing a set of seven theory-based process characteristic of collaborative conjecturing and proving processes. Most centrally, this study provides data on the empirical structure (2.2) of the extracted process characteristics for one exemplary conjecturing and proving task. In addition, this study addresses aspect (2.3) by considering the relatedness of students working collaboratively within one dyad on this conjecturing and proving task. Consequently, study II aims to answer the following research questions: (2.1) Can we find reliable process characteristics for the individual-mathematical and social-discursive component of conjecturing and proving? (2.2) How are individual- mathematical and social-discursive process characteristics of conjecturing and proving interrelated? Can two dimensions of conjecturing and proving processes, one related to the individual-mathematical and one related to the social-discursive component of mathematical argumentation, empirically distinguished from one another? (2.3) Do students working collaboratively on a conjecturing and proving task contribute equally to the quality of conjecturing and proving processes? What is the adequate level of analysis (individual vs. group) for investigating collaborative conjecturing and proving processes?

Study III (Chapter 8) approaches research issue (3) by analysing the relationships between the process characteristics of conjecturing and proving, the quality of the resulting product, and the learning prerequisites (operationalized as prior knowledge on proof). In this study, we apply the analytic framework that we have introduced in chapter 3 and strive to answer the following research questions: (3.1) Which component and process characteristics of collaborative conjecturing and proving are predictive for the quality of the resulting product? (3.2) Do prior knowledge on proof affect the quality of conjecturing and proving processes? (3.3) Do individual-mathematical process characteristics mediate the impact of prior proof knowledge on proof on the final product?

This dissertation concludes (Chapter 9) by comparing and summarizing the findings of the three studies. Limitations and strengths as well as implications for further research projects were discussed. Last, all the findings will be considered in light of some practical implications.

5 Study I

What are process characteristics of successful conjecturing and proving processes? Common claims and findings in research

5.1 Abstract

Identifying the processes that are needed to construct proofs and understanding the actions of those who are (finally) successful in conjecturing and proving are essential to foster students' mathematical argumentation skills. Even though proof and proving are regarded as important in all phases of mathematics education, it seems that there exists more literature on the deficiencies students have than on those characteristics of conjecturing and proving that relate to the successful generation of conjectures and the construction of proofs. The purpose of this literature review is to analyse, categorize, and synthesize past research on conjecturing and proving under these considerations. First, different research perspectives on conjecturing and proving processes as well as the methodological orientation of the underlying studies are presented. We summarize claims and findings of the most representative articles on each research perspective, and analyse the processes that are assumed or reported to be crucial for success. Second, we propose a set of process characteristics that can be considered as indicators of successful conjecturing and proving processes from a theoretical and sometimes even an empirical point of view. Based on the theoretical integration and categorization of findings, we suggest directions for future investigations and practical implications.

5.2 Introduction

A substantial amount of research studies have documented that constructing proofs remains a persistent difficulty for students, even at the tertiary level (e.g., Moore, 1994; Selden, A. & Selden, 2013b). Results of these studies have indicated that students make a variety of mistakes in attempting to construct proofs, including conceptual, logical, formal, as well as strategic errors (e.g., Selden, A. & Selden, 2008; Selden, A. & Selden, 2011). These difficulties may lead students to deduce invalid inferences, use mathematical notations in incorrect ways, or leave students unable to understand the concepts and definitions related to the proving problems. Discovering the deficiencies students have and developing theory-based explanations about how each of these errors could be prevented, can help to improve the teaching and learning of proof.

Another approach that may facilitate the development of specific instructional methods and tools to foster students' mathematical argumentation skills is expanding the knowledge about which processes have shown to be promising in constructing proofs. Thus, carefully analysing the behaviour of those, who are successful in proving, could provide valuable insights into how conjectures may be generated and (deductively) justified (Zazkis et al., 2015). Exposing students to the processes that high-performing students use to construct proofs and that have been recognized to be crucial in various studies, as well as conveying them what can be done to recover from proving impasses (cf. Savic, 2015a), may encourage students to overcome their difficulties.

Over the last years, research on conjecturing and proving has matured and the size of according literature is growing. Therefore, we see the need for systematizing the existing knowledge base on conjecturing and proving processes. The main purpose of this review is to identify common claims and findings about what constitutes “successful” conjecturing and proving processes. We summarize the results of previous studies that describe how and with what intentions specific proving processes are chosen, and analyse the studies’ recommendations on which process characteristics should be taught to students (in order to support their conjecturing and proving skills). It is not expected to find an exhaustive set of process characteristics that covers all processes that may lead to a successful outcome, since the term “successful” in the context of conjecturing and proving may vary between different (research) perspectives.

However, we assume that we are able to identify some more general activities within conjecturing and proving processes, such as exploring the problem situation and identifying appropriate arguments (cf. Boero, 1999) that are considered as necessary for the generation of interesting conjectures and valid proofs from different research perspectives. We categorize these activities that we call *sub-goals within conjecturing and proving processes*, and refer to the single processes, procedures, and operations (*process characteristics*) that are reported as being helpful in achieving these sub-goals. Our aim is to address multiple process characteristics together with their intended goals in order to better understand how these aspects of conjecturing and proving may be related to each other and contribute to a successful outcome (e.g., a conjecture that goes well beyond replicating the premise (cf. Koedinger, 1998) and a sustainable justification that explicitly accounts for why the produced conjecture must be true (Stylianides, A. J., 2007).

In the following, we start with describing the research background before we present our systematic literature review and our analyses.

5.3 Systematizing research perspectives on conjecturing and proving

Conjecturing and proving are complex mathematical processes that comprise different cognitive (e.g., logical, conceptual and problem-solving) as well as social facets (e.g., Weber, 2005). Both activities can be viewed as a particular type of argumentation in mathematics (Pedemonte, 2007). Stylianides, G. J. et al. (2017) distinguished “three broad research perspectives in the area of proof” (p. 244), namely, proving as problem solving, proving as a means of convincing oneself and others of the truth of a conjecture, and proving as a social activity. Within the problem-solving perspective, the generation of a conjecture and the construction of a proof can be considered as tasks in which the learner is asked to find a pattern, to formulate a conjecture, and to find an appropriate justification for it, without initially knowing how to do so (Selden, A. & Selden, 2013a; Weber, 2004). Definitions, already accepted theorems, and acceptable rules of inferences have to be applied to come from the givens to the conclusions, and to evaluate the correctness of a proof. Researchers that conceptualize proving as problem solving are interested in investigating the cognitive and affect-motivational resources, strategies, and skills that are needed to successfully generate a justification that demonstrates that a given statement is true (Stylianides, G. J. et al., 2017).

The illustration of how students think about what constitutes a convincing mathematical argument represents the proving as convincing perspective. Harel and Sowder (1998) categorized students' beliefs about what could remove or create doubts about the truth of an assertion by inducing the notion of proof schemes. Their proof schemes taxonomy distinguished between external, empirical, and analytical proof schemes. From this perspective, the nature, values, and norms of proofs (e.g., Dawkins & Weber, 2017) play an important role. For instance, students' beliefs that the formal-symbolic appearance of a proof is an indispensable requirement for its acceptance may influence their proof-related behaviour (e.g., Harel & Sowder, 1998). These students may prefer to construct proofs that are based on strategies such as "unwrapping definitions" and "pushing symbols" (Furinghetti & Morselli, 2009). The research, which refers to the proving as convincing perspective, is less concerned with the question of how valid and acceptable evidence can be developed, but provides an analytic framework for researchers and educators to examine students' norms related to conviction. In contrast to the problem-solving perspective, which focuses primarily on processes, the convincing perspective considers proof as a product (Stylianides, G. J. et al., 2017).

Treating conjecturing and proving as social activities leads to the view that collaborative argumentative settings provide excellent opportunities to discuss various types of arguments (Yopp & Ely, 2016) and to debate on whether a constructed sequence of arguments constitutes an acceptable mathematical proof (Alibert & Thomas, 1991). Since conjecturing and proving within a social context are tools to communicate and generate mathematical knowledge (e.g., Harel & Sowder, 2007), analysing the studies that have focussed on how students or mathematicians successfully engage in proving to justify or explain a claim to their peers and to communicate their proving ideas, may give interesting insights into conjecturing and proving processes.

Also Stylianides, G. J. et al. (2017), who used these three perspectives of proving as an organizing structure for their research review, remarked that they identified studies that do not fully fit within one of these perspectives. Thus, there might be a broader range of perspectives on proving that better captures the variety of studies that have been conducted in the past. In this contribution, we aim to systematize empirical and theoretical research on conjecturing and proving processes into “research topics” that share a common perspective on these processes. Moreover, we will analyse whether studies from different research perspectives on proving come to different conclusions about which conjecturing and proving processes are needed for the successful generation of conjectures and proofs, and what characterizes promising proving processes (promising in the sense that these processes might be helpful to achieve different sub-goals within conjecturing and proving processes).

5.4 Proving processes

In the past, different models have been established that outline problem solving or, more specifically, conjecturing and proving processes. Models that refer to problem solving usually include the phases of understanding the problem, developing a plan, implementing this plan, and looking back (cf. Carlson & Bloom, 2005; Polya, 1945). The frameworks that deal with proof construction processes additionally emphasize phases that aim to explore the problem situation and possible conjectures as well as phases of systematization (e.g., Boero, 1999; Schwarz et al., 2010). These models are based on self-observations (e.g., Polya, 1945) or emerged from the close observation of other mathematicians (e.g., Carlson & Bloom, 2005; Schwarz et al., 2010). Even though these frameworks have been developed with the intention that (beginning) students may try to implement the described phases and processes, they primarily illustrate how (expert) mathematicians solve problems and write proofs.

Other researchers have also examined the proof-writing behaviour of mathematicians: For instance, the study by Lockwood, Ellis, and Lynch (2016) demonstrated the various ways in which mathematicians used examples to construct proofs. Assuming a continuity between conjecturing and proving, several researchers attributed particular potential to proving processes that are based on informal representations (e.g., Garuti et al., 1998; Sandefur, Mason, Stylianides, & Watson, 2013) or on abductive reasoning (Pedemonte, 2008). However, not all students seem to be able to translate their informal insights into a formal-symbolic (Zazkis et al., 2016) or to link the meaning of objects they have used in their abductive argumentations with the meaning of according objects in their deductive proofs (Pedemonte, 2008).

Using an expert-novice research paradigm, the study by Weber (2001) has shown that successful mathematicians possess a large amount of strategic knowledge that undergraduate students appeared to lack. This knowledge allowed the expert mathematicians to choose adequate proof techniques according to the current situation. From this point of view, it remains an open question to what extent students who have less experience with proof construction are able to employ the same processes that experienced mathematicians usually apply (cf. Reif, 2008).

Another series of studies have started to investigate the approaches of students who were successful in constructing proofs. Conducting task-based interviews with undergraduate students, Gibson (1998) examined how they used diagrams to become familiar with novel problems, to estimate the truth of statements, to develop proving ideas, and to communicate these ideas. The study by Zazkis et al. (2015) analysed the proving behaviour of six highly-successful mathematics majors. They observed a substantial variation in the strategies the students used to construct proofs. Two main strategies (the targeted/ shotgun strategy) were distinguished with regard to how plans were chosen in an attempt to find a solution.

Overall, literature suggests several ways of generating conjectures and constructing proof. However, a common finding in the literature is that successful mathematical problem solvers spend at least some time to think about which rules and theorems are likely useful to apply, and whether they should or should not try to prove a theorem by, for instance, manipulating symbols (e.g., Weber, 2001; Zazkis et al., 2015). In the past, numerous studies have identified key traits and techniques that individuals exhibit while solving proof-related tasks (e.g., Sandefur et al., 2013; Weber, 2004). However, to design adequate learning environments and to support students' learning of proof, we need to know more than just "what works" - we also need to know how and why a specific proving process should be employed. In this research review, we aim to systematize the key traits and techniques that have been described in several research studies and that we call process characteristics of conjecturing and proving in terms of their intended goals.

5.5 The present study

The overall purpose of this review is to analyse the scientific discourse regarding characteristics of successful mathematical conjecturing and proving processes. Based on 126 articles and research reports from 1976 to 2017, we start our investigations by using a statistical clustering method (cf. McCallum, 2002) to describe the research topics and methodological orientation of the included studies. Following the approach by Stylianides, G. J. et al. (2017), we use the identified research topics, which constitute different perspectives on conjecturing and proving, to structure our review. Another goal of this research review is to summarize common claims and findings about aspects of the proving process that are proposed to be crucial for a high proving performance and the learning of proof. Therefore, we analysed the most representative articles on each identified topic (in total 45 articles and research reports) with a specific focus on the sub-goals within conjecturing and proving processes that are considered as necessary intermediate steps for the successful generation of conjectures and the construction of proofs from different research perspectives. Having identified the sub-goals within conjecturing and proving processes, we explore the extent to which each of these sub-goals has been researched so far in the field of mathematics education. Furthermore, we are interested in the processes that are assumed or reported as being helpful in achieving these sub-goals. These process characteristics are discussed in relation to their intended goals in order to better understand how and why students or mathematicians (successfully) employ a specific process.

Particularly, we aim to answer the following research questions:

1) Which research topics within the literature on conjecturing and proving can be identified that represent common perspectives on these processes? Do researchers discuss conjecturing and proving processes from the proving as problem-solving, proving as convincing, or proving as a social activity perspective (as suggested by Stylianides et al. 2017)? What further perspectives on conjecturing and proving processes can be found?

2) Which processes are assumed to be relevant for successful proving performance? What are common claims and descriptions about how mathematicians, mathematics majors, or high-achieving college students construct proofs successfully?

In particular, we are interested in claims about how, and for what purposes, specific proving processes are employed:

3) Which of these proving processes are reported as more general sub-goals within conjecturing and proving? And which process characteristics are considered as being helpful in achieving these sub-goals?

We explicitly do not strive to apply any kind of meta-analytical procedure based of effect-size parameters, as we expect that many studies on conjecturing and proving do not report quantitative findings. Furthermore, the studies as well as their conceptualizations and methods will most likely be too heterogeneous to allow such a statistical summary.

5.6 Method

5.6.1 Literature search

To ensure a systematic approach, we followed a two-step-procedure to select articles and research reports for our review: First, we conducted an extensive database search in MathEduc, ERIC, and ScienceDirect for the keywords “proving” and “proof”. As these three searches resulted in a total of 964533 hits, we decided to restrict our searches by using more specific keywords in each of the three databases. We searched the MatheEduc collection for all journal articles and research reports, which contained the keywords “proof construction”, “proof production”, “proof writing”, “proving activities”, “proving strategy”, “proving strategies”, “proving process”, “proving processes”, “successful proving” as well as “conjecturing” and “deductive reasoning”. For ERIC and ScienceDirect “mathematics” was added as a second key word by using the logical connection “AND”. Regarding the publication year and sources, no restrictions were made.

In a second step, we selected articles and research reports based on their abstracts and the following criteria: Research reports and articles that …

1) … have been published in peer reviewed journals or proceedings;
2) … report to focus on high-quality proving processes or on processes that are strongly related to the production of conjectures and its proofs such as example generation;
3) … contain theoretical claims or empirical findings about successful proving processes or about how, and for what purposes, successful university mathematics students, mathematicians, or high-achieving college students choose a specific process.

5.6.2 Topic modeling

Topic modeling is a method that allows to identify the research perspectives or, respectively, topics that are present in a large collection of documents. It was applied to systematize our selected literature on conjecturing and proving. In this method, a "topic" comprises a cluster of words that usually occur together "in statistically meaningful ways" (Graham, Weingart, & Milligan, 2012, p. 3) and is specified by a probability distribution over the respective words. Inglis and Foster (2018) illustrated topic modeling as a method that builds upon the hypothesis that any given document is a mixture of topics, and thus each document consists of words from possible bags of words, with each bag representing a specific topic.

If this "bag-of-words-assumption" (Steyvers & Griffiths, 2007, p. 427) holds, it becomes possible to mathematically parse a written text into the probable bags, and to define a corresponding probability distribution. Even though topic modeling programs do not have any knowledge about the meaning of words, ignore the order of words, and skip topic-independent words such as "the" and "a", they can be used on a corpus of literature to identify the topics embedded in these documents, without analysing and reading them individually (Graham et al., 2012; Inglis & Foster, 2018).

We followed the approach proposed by Inglis and Foster (2018). All of our articles and research reports were stored as pdf-files and converted into .txt-files using ABBYY FineReader OCR Pro. Copyright statements as well as information about the journal were removed as they do not contribute to the content the documents contain. For our analysis, we used the command-line topic-modeling program MALLET (version 2.0.8RC2; McCallum, 2002). This program requires that the number of topics that the algorithm should detect has to be set by the user. Therefore, the "perplexity" of a model with a given number of topics was calculated by running the topic model program to a subset of the selected documents and evaluating the resulting fit parameters. We went through this process repeatedly by systematically changing the number of topics. Low perplexity values indicate good model fits. Increasing the number of topics leads to the reduction of the perplexity, but at some point the interpretation of too many topics becomes challenging. Thus, it is suggested to use a method comparable to Catells's scree test

(1966) for exploratory factor analysis. Inglis and Foster (2018, p. 477) propose to find “a point at which the reduction in perplexity appears to ‘level off’” and to select this point as the number of topics the algorithm should identify. Once the number of topics has been determined, the program was run again in order to return a defining list of words composing those topics. Regarding research question (1), we interpreted each topic by analysing the defining list of words and by studying the most representative articles or research reports on each topic (the papers with the highest proportion of words from their corresponding topic). Then, we tried to find meaningful descriptions for each of these topics and allocated these topics to the research perspectives on proving as outlined by Stylianides, G. J. et al. (2017). The created topic model served as an organizing structure for the further analyses and helped us to reduce the complexity of our unstructured document collection of literature on conjecturing and proving. Since we would not have been able to handle the deluge of data that would have resulted if we had tried to read and interpret all the relevant literature on conjecture and proving, we have limited the following in-depth analysis to 45 (of 51) articles that represent the range of topics and perspectives found in the studies.

5.6.3 In-depth-analysis

In the in-depth analysis, we included the three most representative articles or research reports, for each of the identified topics. First, we highlighted their claims and empirical findings about any type of promising conjecturing and proving processes. These data were analysed using “the synthesis of qualitative research approach” outlined by J. Thomas and Harden (2008) to enhance transparency in the review process. Due to the fact that it is difficult to deal with the question of what counts as data or - with regard to our review - as claim or finding when analysing quantitative and especially qualitative research, we curbed this problem by focusing mainly on the listed claims and findings in the text sections labelled as “results”, “findings”, “discussion”, or “implications for teaching and learning”. Claims and findings in the abstracts or in other text sections that appeared to be relevant as well and that were reported in a similar way as in the text sections labelled as “results”, “findings”, “discussion”, or “implications for teaching and learning” were also taken into account. In general, when deciding whether to include a claim or finding in our analysis or not, we relied on the second and third criteria that we already had defined and used when searching for relevant literature on conjecturing and proving.

The synthesis approach proposed by Thomas, J. and Harden (2008) comprises three steps that we adapted in the following way: (i) Free line-by-line coding of the claims and empirical findings from the primary studies: The extracted data were analysed for meaning and content during the coding. The developed codes were subsequently listed into a code book. This procedure enabled us to translate the codes and concepts between the studies; (ii) Finding descriptive categories: The developed codes were examined for their meanings, and reorganized into related categories; (iii) Generating analytical categories to directly address the second and third research question: Each category was examined and compared to other categories, in particular by searching for similarities and differences. The categories were analysed by whether and how they report any theoretical relationship or empirical mechanism between proving behaviour, and proving performance. Similar categories were merged into higher-level, more abstract categories that sometimes went beyond the claims and findings of the original studies. The step of going beyond the claims and findings of the original studies incorporated, for example, our interpretation of which processes were rather described as sub-goals within conjecturing and proving or as characteristics that might be helpful in achieving these sub-goals. This step of the analysis led to a structured description of which proving processes were considered as relevant for the successful generation of conjectures and proofs.

The first two steps were mainly inductively. In the third step, most of the categories were derived from the data, but we already had some categories in mind before we generated the more abstract categories and therefore followed an approach that is partly inductive and partly deductive. The more abstract categories we already had in mind were based on the phases described in the existing frameworks of proof construction (cf. Boero, 1999; Schwarz et al., 2010) or problem solving (cf. Carlson & Bloom, 2005; Polya, 1945). These pre-defined categories, which we modified and refined throughout our in-depth analysis, were: exploring the problem situation, organizing single inferences into a chain, communicating arguments, and transforming informal into formal arguments as well as using informal representations, generating examples, using formal symbols, or metacognitive processes. Besides these categories, we were interested in finding new categories that were described either as sub-goals within conjecturing and proving or as process characteristics that might be helpful in achieving one or more of these sub-goals.

Table 1 sets out the dimensions that structure our in-depth analysis. An example of how claims or empirical findings were coded and categorized is given in Table 2.

Table 1: *Dimensions that frame the in-depth analysis of claims and findings about promising conjecturing and proving processes*

	Description
Type of sub-goals	The aim of this dimension is to identify the different sub-goals within conjecturing and proving processes that have been researched in the field of mathematics education. This dimension about the sub-goals within conjecturing and proving processes comprises the intermediate steps that have been considered as necessary or, at least, as central for the successful generation of conjectures and the construction of proofs from different research perspectives. It addresses the reasons for choosing a specific proving process or procedure. Processes that have been described as sub-goals within conjecturing and proving might be placed in categories such as exploring the problem situation, organizing single inferences into a chain, communicating arguments, transforming informal into formal arguments, … It might be that in some studies the sub-goals were not described in detail and more as latent constructs, as if their meaning were obvious. In these cases, the sub-goals were also taken into account.
Type of process characteristics	The aim of this dimension is to identify the different processes that have been considered (or at least can be interpreted as such by taking the context of the study into account) as being helpful in achieving one or more of these sub-goals within conjecturing and proving processes. It provides the basis for understanding how the use of a particular proving process or procedure can work and thus contribute to a successful outcome. Processes that have been described as process characteristics might be placed in categories such as example generation, using informal representation systems, using formal symbols, meta-cognitive processes, … The process characteristics are recognizable by the fact that they can be operationalized for assessing purposes or discussed with students in a classroom environment to help alleviate difficulties in proving tasks.

Table 2: *Coding examples of claims and findings from three studies. The study of by Savic (2015a) represents the topic "problem-solving with a specific focus on processes, impasses, and incubation", the study of by Martinez, Brizuela, and Superfine (2011) represents the topic "modeling", and the study of by Blanton and Stylianou (2014) represents the topic "social/ collective argumentation".*

Claims and findings (data)	Free-line-coding	Sub-goals	Process-characteristics
"Dr. B's actions to overcome his impasse included moving on to the next theorem, creating counterexamples, and being interrupted by his family, where, at lunch, he had an insight that turned out to be useful for furthering his proof." (p. 74)	processes to overcome impasses, including incubation; counterexamples, moving on, doing something else	recovering from impasses	generating examples: trying to create a counterexample incubation strategies in form of domain-general problem-solving strategies: moving on to the next task; resting and "sleeping on it"
"Our intention was that in producing a chain of equivalent expressions, students would use one of the aspects of algebra […] to make explicit something that was implicit in the initial algebraic expression" (p. 36)	manipulating formal expressions, making something explicit, searching for a new expression	exploration - finding an adequate representation for the proof	using and manipulating formal-symbolic/ algebraic representations
"… the number of student transactive utterances increased from 27% to 64%, suggests that they had come to view proving as a habit of mind that involved explaining, critiquing, justifying, and so forth, as the means to […] negotiate meanings for the component parts of a conjecture (e.g., "center" of a group)." (p. 96)	transactive processes (explaining, critiquing, justifying), negotiating meanings	generating a shared understanding	transactive processes: explaining and critiquing

5.7 Results

5.7.1 Main topics of the literature in the area of proof construction

Our first research question addressed the systemization of the literature on conjecturing and proving into research topics that represent common perspectives on these processes. In particular, we were interested in identifying and refining the three research perspectives on proving (proving as problem-solving, proving as convincing, or proving as a social activity) that have been suggested by Stylianides, G. J. et al. (2017). To determine the optimal number of topics for out topic modelling analysis, we used the perplexity of models with 5 to 35 topics. The perplexity graph showed that a model with 17 topics was the most adequate approach for presenting the literature on conjecturing and proving (Figure 7).

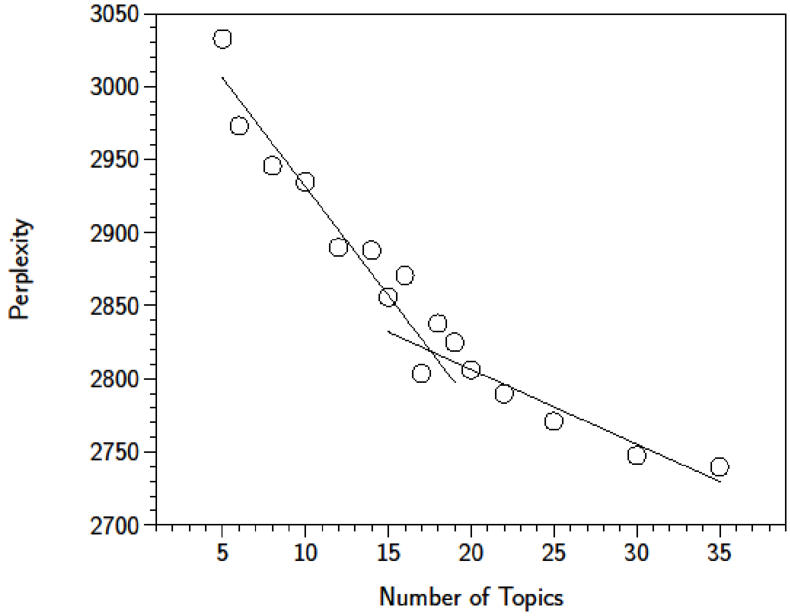

Figure 7: The perplexity of topic models with varying numbers of topics (5 to 35 topics in jumps of 1 or 2). The two lines represent linear approximations for the perplexity level above and below the identified point 17, where the graph 'levels off'.

The research topics determined by the algorithm, which serve as an organizing structure for our research review, are listed in Table 3. These are the topics that would be most likely to have led to the articles and research reports of our document collection on conjecturing and proving. Table 3 also shows the names and perspectives that we allocated to the topics, the defining list of words, as well as the average proportion of documents assigned to the topics.

In our collection of documents, we were able to detect all the three perspectives on proving outlined by Stylianides, G. J. et al. (2017). Six of the identified topics were related to the problem solving perspective, namely the "examples and conjecturing" topic, "the problem solving with a specific focus on affects" topic, the "thinking processes" topic, the "informal understanding" topic, the "problem solving with a specific focus on processes, impasses, and incubation", and the "types of reasoning (e.g., semantic/syntactic)" topic. Articles and research reports that represented "the problem solving with a specific focus on affects" topic and "problem solving with a specific focus on processes, impasses, and incubation" topic were explicitly based on the problem-solving perspective. The documents that were assigned to the “examples and conjecturing” topic, “informal understanding” topic, or the “types of reasoning (e.g., semantic/syntactic)” topic were implicitly linked to the problem-solving perspective as they referred to the development and use of specific problem-solving strategies. Since problem-solving tasks usually require complex thinking processes and therefore the use of heuristic strategies that can, but do not necessarily, lead to a solution (Abel, 2003), it appeared obvious to allocate these research topics to the problem-solving perspective as well.

The documents that mainly represented the “university level” topic or the “school students” topic shared the convincing perspective on proof and, additionally, characterized proving as an educational learning goal, which is central within different educational levels. They categorized and described the different types of processes that school students, undergraduates, or mathematicians used to construct proofs and to convince themselves or the corresponding mathematical community. The proving as a social activity perspective was covered by articles and research reports that had a high proportion of words from the “social/ collective argumentation” topic. Within this perspective, proving was conceptualized as an activity that occurs in a social context. The documents that were clustered in one of the topics that shared one of the three entrenched perspectives on proving have explicitly or implicitly treated proving as a problem-solving, convincing, or a socially-embedded activity at the expense of emphasizing the discovery function of conjecture generation and proof construction. Both, the “geometry (conjecturing)” topic and the “modeling” topic were about proof-related activities that could be used to generate new mathematical knowledge (within a specific content area) and thus, pointed to the discovery function of proof. The perspective that proof and proof-related activities serve as means to systematize mathematical knowledge was taken up (besides other perspectives) in the literature that represented the “formal system” topic. We were not able to find further perspectives on conjecturing and proving in the included documents.

Five topics, “the nature of proof and teaching of proof” topic, the “argumentation structure” topic, the “geometry (proving)” topic, the “formal systems” topic, as well as the “abstract and linear algebra” topic, could not be clearly attributed to one specific perspective on conjecturing and proving, since the articles and research reports with the highest proportions of words from (at least) one of these topics contributed to the conceptualization of conjecturing and proving from different perspectives. The “nature of proof and teaching of proof” topic captured the literature on proving that provided a global view on proof-related activities by discussing the goals and situations that may guide each of these activities. Although the topic could not be attributed to one of the perspectives, the convincing perspective appeared to play an important role within this topic. Regarding the “argument structure” topic, we observed that articles and research reports that referred to this topic were concerned with the cognitive and structural continuities and distances between argumentation and proof. This research issue seemed to be clearly related to the problem-solving as well as the discovery perspective. In the literature that was most representative for the "geometry (proving)" topic, proof-related activities were discussed with respect to the conceptual (content-specific) knowledge that is required to achieve success on geometric proving tasks. From this point of view, it could be assigned to the problem-solving perspective, but we have also identified the discovery and social activity perspective within this kind of research. Articles and research reports with a high proportion of words from the "formal system" topic took, as already mentioned, proving as a means for systematizing knowledge perspective as well as the discovery perspective into account. The "abstraction and linear algebra" topic included literature on proving, which focused on students' knowledge about different linear algebra concepts, as well as literature that discussed methods of theorem proving with abstraction. The latter referred to "the automatic theorem proving" topic, a topic that has been identified as an independent branch of research. Due to the fact that the literature related to the "abstraction and linear algebra" topic differed widely and thus had no common theme, we were not able to assign a research perspective to this topic. The articles and research reports on "automatic theorem proving" were excluded from further analyses, as the focus of this literature review was on the mathematics educational literature on conjecture and proving.

The topic modeling analysis showed that the included documents typically consisted predominantly of words from topics with an overarching perspective on conjecturing and proving, followed by words from topics with (in descending order) the problem-solving perspective, the convincing perspective, the discovery perspective, and finally the perspective on proving as a socially-embedded activity. Within the problem-solving perspective the "example and conjecturing" topic appeared to be most popular, followed by the "problem-solving with a specific focus on affects" topic. Regarding the topics that shared the convincing perspective, the algorithm showed that words from the "university level" topic occurred slightly more frequently than words from the "school students and school teacher" topic. Overall, the literature on conjecturing and proving was composed of only a small proportion of words from the "social/ collective argumentation" topic ($M_{\text{proportion}} = 0.026$) and thus only a few articles explicitly shared the social activity perspective on proving. Most articles and research reports seemed to consist of words from the "nature of proof and teaching of proof" topic ($M_{\text{proportion}} = 0.363$), typifying the "overarching perspective on proof".

Table 3: *The 17 topics and the defining list of words (the words that best characterize the corresponding topic in order of probability) for each topic (sorted by their average proportion).*

Perspective	Name of the topic	Similarities	Defining list of words	Average proportion
overarching (mostly convincing)	nature of proof and teaching of proof	Identifying and describing different proof-related activities with respect to their situations in which they are embedded, their specific goals and functions.	proof mathematics mathematical students research education proving study reasoning teaching proofs knowledge prove student learning analysis journal university understanding case	0.363
convincing perspective – focusing on a specific educational level	university level	Several approaches that professional mathematicians, undergraduate, or doctoral students use to prove theorems based on the ideas they find convincing or (intuitively) meaningful.	students proof proofs weber participants statement undergraduates courses group alcock writing asked mathematics prove student undergraduate mathematicians strategies it's statements	0.085
convincing perspective – focusing on a specific educational level	school students and teachers	Assessing school students' and school teachers' proof performance by focusing on the difficulties they encountered, the proof schemes they used, and their understanding of proofs and refutations.	proof students teachers number mathematics mathematical arguments reasoning argument numbers deductive proofs set schemes prime education statement proposition study prospective	0.071
problem-solving perspective – focusing on strategy use	examples and conjecturing	Using examples as a powerful tool to explore the problem situation, to estimate the truth and to justify conjectures.	examples students conjecture numbers number conjectures true consecutive work strategies mathematicians general task case proving cases stylianides insight activity multiple	0.067
discovery perspective – focusing on a specific content area	geometry (conjecturing)	Generating new knowledge (e.g., statements or conjectures) in the context of geometry by	students fig conjecture case geometry conjecturing proof proofs conjectures deductive point tasks design triangles argumentation teacher	0.058

		using (counter-) examples or by reflecting on and utilizing already constructed proofs.	triangle function activity statement	
problem-solving perspective – focusing on affective and cognitive resources	problem solving (especially, affect)	Studying the interrelation between affective and cognitive resources and how they influence one's problem-solving behaviour.	problem students solving proving mathematics induction problems solution cognitive university process problem-solving activity representation work secondary beliefs behavior studies solve	0.050
overarching (mostly discovery and problem-solving)	argumentation structure	Analysing the cognitive continuities and structural distances between argumentation and proof.	argumentation proof claim students argument model abduction number data case algebraic arguments cases rule conjecture structure claims arithmetic warrant generalization	0.041
problem-solving perspective – focusing on strategy use	thinking processes	Abstraction and creative thinking as central components of conjecturing and proving processes.	problem students process solving geometry knowledge student investigation thinking learning problems test ability processes mathematical model posing problem-solving figure metacognitive	0.039
problem-solving perspective – focusing on strategy use	informal understanding	Analysing the cognitive processes that are involved in concept image-, visual intuitions- and example generation activities.	function image proof concept argument graphical process arguments informal formal graph definition springer processes interval diagram diagrams generation participants derivative	0.035
overarching (mostly problem solving) - focusing on a specific content/ conceptual knowledge	geometry (proving)	Prior knowledge and activities related to geometry proof construction.	triangle concept definition activity group students triangles definitions angles angle line formal perpendicular development image properties thinking individual geometry point	0.033
problem-solving perspective – focusing on approaches to	problem solving (esp. processes, impasses, incubation)	The processes mathematicians take to recover	theorem proving pupils mathematicians problem process selden phase actions identity construction time	0.032

overcome proving impasses		from proving impasses.	incubation element behavioral planning solving impasses framework mind	
social-activity perspective	social/ collective argumentation	Using collaboration as a resource for building mathematical arguments and producing understanding.	students discourse teacher transactive argument authority arguments utterances teachers episode classroom learning ideas reasoning episodes finley empirical collective odd mathematical	0.026
problem-solving perspective – focusing on strategy use	types of reasoning (e.g., semantic/ syntactic)	Semantic and syntactic proof production and other representation systems that best model students' reasoning.	reasoning theory lines semantic indirect students line theorem kirk formal statement meaning meanings system proof points syntactic contradiction mathematical point	0.025
overarching – (mostly discovery and systematization)	formal systems	Discovering and proving theorems by the application of (logical) rules, meta-rules, and formal symbols.	rules rule knowledge theorems system image theorem proved prove set element item base properties neighbourhood mathematicians general property sets found	0.023
discovery perspective – focusing on specific content areas	modeling (with algebraic expressions)	Algebra as a modeling tool for formulating conjectures, solving equations, and generating proofs within different content areas by defining and using variables and parameters.	students blend algebra episode blending efp ppp variables problem fig expression geometric i.e conceptual line kgb outcome equation data Stacey	0.020
(no assignment to a perspective)	automated theorem proving	Related to the field of machine learning.	transformation facts proofs proof transformations fact mash isabelle set transfer springer number theory mizar provers sledgehammer mepo type theorem goal	0.017
overarching – not clearly assignable to one or more perspectives	abstraction and linear algebra	Prior knowledge and activities related to theorem proving	proof abstraction linear clauses set search clause proofs span suppose resolution strategies depth	0.016

	in the context of linear algebra or about abstraction methods to develop automated theorem proving systems.	abstractions vectors vector theorem matrix strategy m-clauses

5.7.2 Methodological orientation of the underlying studies

In the following, we describe the methodological orientations of the articles and research reports that are composed of particularly high proportions of words from one of the identified topics. We included a total of 51 papers in the search for central claims and empirical findings about promising proving processes, and finally 45 of these references comprised relevant claims or empirical findings about sub-goals within conjecturing and proving processes or about the process characteristics that may be helpful in achieving these goals. The selected articles and research reports were published between 1976 and 2017, with a median publishing year of 2013 and mean of 2011. The majority of the papers used small sample sizes (*Median* = 8.5). One paper did not provide any sample size information and eight papers contained no empirical data at all. Regarding the studies' methods, mainly qualitative approaches were applied. In the 34 qualitative studies (excluding the 8 theoretical papers and 3 quantitative ones), real-time data collection techniques such as task-based interviews or live scribe pens, in which the participants were asked to solve the problem aloud, were most prevalent. The resulting think-aloud-protocols served as data base for the subsequent analyses. Other common data collection techniques to make proving strategies and processes accessible were questionnaires, course observations, and collaborative settings. In three studies, the participants were interviewed only after they had completed the given tasks. They were prompted to recall the thoughts they had in mind while solving the proof problems and retrospective-think-aloud-protocols were created.

Concerning the data analysis, most researchers developed open codes to describe their data in the spirit of grounded theory (e.g., Mueller et al., 2012; Uğurel, Moralı, Koyunkaya, & Karahan, 2016). In the remaining three studies, quantitative research was conducted and corresponding statistical methods were applied. Two of these research papers reported the results of experimental design studies that were implemented to investigate the influence of specific instructional approaches on the learning and teaching of proof.

To sum up, we observed variations in the studies' design principles (e.g., time-limitations vs. no time-limitations, intervening or not intervening the solution process by instructors), in the domains (e.g., geometry, number theory, analysis), in the type of tasks (e.g., open-ended, similar to already known tasks), in the populations (mathematicians, PhD students, (under-)

graduate students, (junior) high-school students), in the data sources (e.g., interview transcripts, video-recordings of collaborative settings, questionnaires) and in the research (theoretical, quantitative, qualitative). Moreover, results showed that the research on conjecturing and proving processes is based mostly on qualitative, explorative studies. Studies that tested hypotheses and made predictions about participants' proving behaviour or performance were rare. Table 4 gives an extended overview of the study characteristics of all selected articles and research reports.

Table 4: *An overview of the 45 papers that constitute our data for the qualitative analyses.*

Name of the topic	Article/ research report	Sample	Method, instruments, analysis
Nature of proof and teaching of proof	Mejía-Ramos and Inglis (2009)	-	Theoretical paper
	Zaslavsky, Nickerson, Stylianides, Kidron, and Winicki-Landman (2012)	-	Theoretical paper
University level	Weber (2004)	14	Qualitative, think aloud, task-related interviews
	Weber and Alcock (2004)	14	Qualitative, description of previous studies: questionnaires, task-related interview, think aloud
	Zazkis et al. (2015)	6	Qualitative, retrospective recall, interpretative analysis
School students and teachers	Lin, F.-L., Yang, and Chen (2004)	3345	Quantitative, questionnaires, theory-based coding
	Uğurel et al. (2016)	15	Qualitative, think aloud, interview, grounded theory
	Lee, K. (2016)	60	Qualitative, written proofs, theory-based coding
Examples and conjecturing	Ellis et al. (2017)	38	Qualitative, task-based interviews, coding was to develop open codes (grounded theory)
	Ozgur, Ellis, Vinsonhaler, Dogan, and Knuth (2017)	38	Qualitative, task-based interviews, theory-based coding

	Ellis, Lockwood, Dogan, and Williams (2013)	26	Qualitative, interviews, think aloud, theory-based coding and open-coding process
Geometry (conjecturing)	Komatsu (2011)	2	Qualitative, questionnaire, observation, interview
	Komatsu, Tsujiyama, and Sakamaki (2014)	4	Qualitative, interviews, case study, protocol analysis
	Komatsu (2016)	4	Qualitative, interviews, case study, protocol analysis
Problem solving (especially, affect)	Furinghetti and Morselli (2004)	1	Qualitative, observations, protocol analysis
	Furinghetti and Morselli (2009)	2	Qualitative, written individual reports, protocol-analysis, interpretation
	Furinghetti, Maggiani, and Morselli (2013)	9	Qualitative, questionnaires, interviews, theory-based coding
Argumentation structure	Pedemonte (2008)	-	Theoretical paper
	Pedemonte and Reid (2011)	-	Theoretical paper
	Pedemonte & Buchbinder (2011)	12	Qualitative, task-based interviews, think aloud, theory-based coding
Thinking processes	Herlina and Batusangkar (2015)	78	Quantitative, experimental design, intervention, interview, inferential statistics
	Huda (2016)	2	Qualitative, written proof, interview, protocol-analysis
	Carroll (1977)	201	Quantitative, experimental design, pre-and post-test, inferential statistics
Informal understanding	Kidron and Dreyfus (2014)	2	Qualitative, interview, protocol-analysis, interpretation
	Antonini (2011)	21	Qualitative, clinical interview (task-based)

	Zazkis and Villanueva (2016)	8	Qualitative, clinical interview (task-based), grounded theory
Geometry (proving)	Zandieh and Rasmussen (2010)	25	Qualitative, video-recordings, interviews, homework and exams, portfolios, retrospective reports
	Vidakovic and Martin (2004)	4	Qualitative, written work, stimulated recall, interview, protocol-analysis, interpretation
	Küchemann and Hoyles (2006)	2	Qualitative, collaborative setting, protocol-analysis
Problem solving (esp. processes impasses, incubation)	Savic (2015b)	2	Qualitative, Livescribe pen data, interviews, theory-based coding
	Savic (2015a)	9	Qualitative, Livescribe pen data, interviews, theory-based coding
	Selden, A. et al. (2010)	No exact number given	Qualitative, design experiment, course observations, interviews
Social/collective argumentation	Mueller et al. (2012)	8	Qualitative, collaborative setting, (inductively based analysis) grounded theory
	Blanton and Stylianou (2014)	30	Qualitative, classroom observations, written proofs, theory-based coding
	Otten et al. (2017)	14	Qualitative, classroom observations, discourse-analysis
Types of reasoning (e.g., semantic/ syntactic)	Dawkins (2012)	2	Qualitative, task-based interviews, protocol-analysis
	Dawkins (2015)	2	Qualitative, task-based interviews, protocol-analysis
	Dawkins and Karunakaran (2016)	-	Theoretical paper
Formal systems	Pastre (1989)	2	Qualitative, task-based interview, mainly: system experiment
	Bagchi and Wells (1998)	-	Theoretical paper

	Grenier (2013)	-	Theoretical paper
Modeling	Winkel (2015)	-	Theoretical paper
	Zandieh, Roh, and Knapp (2014)	4	Qualitative, collaborative setting, videotape recordings, grounded theory
	Martinez et al. (2011)	9	Qualitative, task-related interviews, grounded theory
Linear algebra	Plaxco and Wawro (2015)	5	Qualitative, task-related interviews, grounded theory

Our analysis of claims and empirical findings captured both sub-goals within conjecturing and proving processes as well as the process characteristics that have been regarded as being helpful in achieving one or more of these sub-goals. Below we present the framework (that we have developed out of the results from our in-depth analysis) in two parts: First, we introduce the process-characteristics that we have identified. Second, we address the relationships between the sub-goals and the processes characteristics by presenting the co-occurrences that have been discussed in the literature (or at least could be interpreted as such by taking the context in which the study was embedded into account) between some sub-goals and some process characteristics.

5.7.3 Sub-goals within conjecturing and proving

Based on our research questions, we tried to identify common claims and empirical findings about promising conjecturing and proving processes. We observed that a large number of researchers conceptualized conjecturing and proving as a pathway including several intermediate steps (e.g., Furinghetti & Morselli, 2009; Mejía-Ramos & Inglis, 2009; Savic, 2015b; Uğurel et al., 2016; Weber, 2004; Zazkis et al., 2015). How these steps may be achieved, remained an open question in some of these studies (e.g., Mejía-Ramos & Inglis, 2009). Yet, in other studies, we found out that different processes were mentioned in the sense that they may be useful for reaching one or even more of these intermediate steps (e.g., Ellis et al., 2017; Furinghetti & Morselli, 2009; Weber & Alcock, 2004). This led us to differentiate between the *sub-goals*, these are the intermediate steps within conjecturing and proving processes, and the *process characteristics* that were considered as being helpful for achieving these sub-goals.

We started with the identification of the sub-goals that were associated with the successful generation of conjectures and the construction of proofs. The study by Mejía-Ramos and Inglis

(2009) suggested to distinguish three main processes in the context of proof construction, namely “the exploration of a problem”, “the estimation of truth of a conjecture”, and “the justification of a statement estimated to be true” (p. 3). Moreover, Mejía-Ramos and Inglis (2009) set up the process of presenting an argument as another crucial proof-related activity. These four main processes served as a starting point for detecting further sub-goals within conjecturing and proving processes as well as for refining the existing sub-goals from our document collection. Prior research has shown that in particular the studies that conceptualized proving from a problem-solving perspective emphasized the importance of understanding and exploring the problem situation (Furinghetti & Morselli, 2009; Savic, 2015; Zazkis et al., 2015). The studies that addressed the use of examples additionally pointed to the step of estimating the truth of a conjecture (e.g., Ellis et al., 2017; Ozgur et al., 2017). Accordingly, we decided to divide the sub-goal *exploring the problem situation* into more fine-grained sub-goals, as we have observed that some of the studies focused more on the exploratory sub-goal *developing a strong understanding of the statement to be proved* (e.g., Antonini, 2011; Ellis et al., 2017; Zazkis et al., 2015), other studies more on the sub-goal *inventing and formulating new conjectures or refining existing conjectures* (e.g., Ellis et al., 2017; Komatsu, 2011; Komatsu et al., 2014; Ozgur et al., 2017), and yet others more on the sub-goal *finding an adequate representation for the proof and adequate proving strategy* (e.g., Martinez et al., 2011; Mueller et al., 2012). The sub-goal *developing a strong understanding of the statement to be proved* seemed to be closely linked to the sub-goal *estimation of the truth*. We merged these two sub-goals into one category. The process of *justification* in its various forms has been discussed in articles from almost all topics. Some researchers described *example-based justifications* (e.g., Ellis et al., 2017; Komatsu, 2016; Ozgur et al., 2017), others drew attention to the production of general deductive arguments. Regarding the *production of general deductive arguments*, we observed that both sub-goals *drawing inferences* (e.g., Furinghetti et al., 2013; Weber & Alcock, 2004) and *structuring and organizing inferences* (e.g., Bagchi & Wells, 1998; Pedemonte & Buchbinder, 2011) were considered as crucial steps. In addition, several articles, especially those dealing with the “examples and conjecturing“ topic, the “types of reasoning” topic, or the “informal understanding” topic, discussed the step of *translating less formal to formal arguments,* which we defined as a further sub-goal within conjecturing and proving processes. The sub-goal *communicating and presenting arguments* was mainly described in articles and research reports with a high proportion of words from the “nature of proof and teaching of proof” topic as well from the “formal systems” topic. Articles and research reports representing the “problem solving with a specific focus on processes, impasses, and incubation” topic obviously contributed to the sub-goal *resolving fixations/ avoiding errors*. Yet, this sub-goal was also addressed in articles and research reports from other topics (e.g., Furinghetti & Morselli, 2004; Huda, 2016). Articles and research reports that were clustered to the “social/ collective argumentation” topic considered *generating a shared understanding* as central sub-goal. The last sub-goal that we identified within the literature on conjecturing and proving was *producing understanding about the proof* (e.g., Savic, 2015b; Weber, 2004). Table 5 presents an overview about the sub-goals that we have inferred from the literature on conjecturing and proving.

Table 5: *Definitions of the sub-goal within conjecturing and proving processes derived from 45 articles and research reports on conjecture and proving.*

Sub-goals	Claims and findings that were attended to this sub-goal comprised descriptions about the process-characteristics that might help to …
Developing a strong understanding of the statement to be proved/ estimation of the truth	… become familiar with the concepts and definitions related to the statement to be proved or to gain initial insight into why the conjecture must be true (or false).
Inventing and formulating new conjectures or refining existing conjectures	… find patterns and to identify structures from which conjectures can be inferred or adapting already stated conjectures.
Finding an adequate representation for the proof and an adequate proving strategy	... find a representation that enables to detect permissible configurations and structures and to work with it as well as to contemplate different solution approaches by estimating the effects of different proving strategies and tools.
Generating example-based justifications	… generate counter-examples to refute a conjecture or examples that enable to construct a viable proof.
Drawing inferences	… ensure that the individual constructed arguments are structurally sound.
Structuring and organizing inferences	… ensure that the constructed chain of arguments is structurally sound.
Translating less formal to formal arguments	… to systematize informal arguments and connect them to formal representations and axiomatic arguments (arguments from a shared knowledge base)
Communicating and presenting arguments	… to communicate arguments accurately and precisely and to avoid the presentation of arguments that are open to different or even incorrect interpretations.

Resolving fixations/ avoiding errors	… to develop anticipatory thinking, to recover from impasses, and to rework parts of arguments in the case of a wrong direction.
Generating a shared understanding	… to construct a shared knowledge base that is grounded in facts and to pursue common goals.
Producing understanding of the proof	… to sustain understanding why the argumentation leads to a proof, to understand the causal mechanism behind the proof, and to make sense of the associated representations and concepts.

Our analysis of claims and empirical findings captured both sub-goals within conjecturing and proving processes as well as the process characteristics that have been regarded as being helpful in achieving one or more of these sub-goals. Below we present the framework (that we have developed out of the results from our in-depth analysis) in two parts: First, we introduce the process-characteristics that we have identified. Second, we address the relationships between the sub-goals and the processes characteristics by presenting the co-occurrences that have been discussed in the literature (or at least could be interpreted as such by taking the context in which the study was embedded into account) between some sub-goals and some process characteristics.

5.7.4 Process characteristics of conjecturing and proving

The majority of the process characteristics we extracted from the literature were related to *example use*. Across the body of research on example use, we observed that some researchers described example use more generally (e.g., Antonini, 2011; Furinghetti & Morselli, 2004; Lee, 2016). In contrast, others have identified and categorized different types of examples and studied the relation between the specific types of examples and how learners leverage their thinking with those examples in order to support conjectures, to see similarities, and to generalize arguments. We observed that the following types of example-based processes were distinguished (cf. Ellis et al., 2017; Komatsu, 2011; Ozgur et al., 2017): *varying examples systematically, considering boundary cases, choosing examples with specific properties, testing a diversity of examples,* and *attempting to construct counterexamples*. *Working with informal representations* appeared to be a further category of process characteristics researchers have paid attention to (e.g., Kidron & Dreyfus, 2014; Pastre, 1989). This category includes the use of verbal language (e.g., Furinghetti et al., 2013; Pastre, 1989), diagrams (e.g., Küchemann & Hoyles, 2006; Weber, 2004), and mental pictures (e.g., Kidron

& Dreyfus, 2014). The next category of process characteristics that we identified was termed *using formal symbols and algebraic representations*. Different researchers emphasized the role that formal symbols and algebraic representations could play in order to construct formal-deductive proofs (e.g., Pedemonte & Buchbinder, 2011; Weber & Alcock, 2004). *Reformulating* and *using abductive and inductive inferences* are two categories of process characteristics that were closely linked to the process characteristics of *working with informal representations* and of *using formal symbols and algebraic representations.* The category of reformulating involves the process characteristic of changing between different representation systems (Furinghetti & Morselli, 2009) as well as supplanting informal concepts with formal ones (Dawkins, 2015). Literature that referred to the category of *using abductive and inductive inferences* described, for instance, how abductions that occurred in learners' mathematical activities could be productively used within conjecturing and proving processes (e.g., Pedemonte & Reid, 2011) or how the generation of examples could enable to make generalizations (e.g., Ellis et al., 2017; Pedemonte & Buchbinder, 2011). *Unpacking mathematical statements* was another process characteristics category that we inferred from the literature and that appeared to be strongly related to the category of *using formal symbols and algebraic representations.* It comprised the formulation of the formal-rhetorical part of a proof (Selden, McKee, & Selden, 2010) and the use of the logical structure of statements (e.g., Dawkins, 2012; Zazkis & Villanueva, 2016). *Organizing one's knowledge and existing definitions, concepts, and structures* included all process characteristics that were related to the process of (re-)organizing previous "mathematical constructs within mathematics and by mathematical means so as to lead to a construct that is new to the learner" (Kidron & Dreyfus, 2014, p. 299). This included process characteristics such as collecting and establishing links between one's ideas (Kidron & Dreyfus, 2014) or combining information in novel ways (Zandieh et al., 2014). We observed that a large amount of articles and research reports discussed the role of *meta-cognitive processes* within conjecturing and proving. The claims and findings of these studies underlined the importance of reflecting on one's own work (e.g., Komatsu et al., 2014; Selden, A. et al., 2010) or on one's strategy use (e.g., Zazkis et al., 2015), anticipatory thinking (e.g., Furinghetti & Morselli, 2009), as well as evaluating one's own incorrect proof attempts (Savic, 2015). The next category of process characteristics that we extracted from the literature was *applying domain-general problem-solving strategies.* The process characteristics of *drawing analogies* (e.g., Herlina & Batusangkar, 2015), *splitting the task into sub-tasks* (Pastre, 1989), *working forwards and backwards* (e.g., Carroll, 1977), and *applying the trials-and-error strategy* (e.g., Zazkis et al., 2015) were allocated to this category. *Transactive activities* such as *taking the arguments of others into account* (Mueller et al., 2012) and *externalizing one's own ideas* (Vidakovic & Martin, 2004) represented another category of process characteristics that we inferred from the claims and findings that constituted our data base.

5.7.5 The interaction between sub-goals and process characteristics

The development of our framework of “sub-goals versus process characteristics” (see Table 6) was guided by the following questions: What are the assumptions about how to achieve the sub-goals within conjecturing and proving processes? Which process characteristics have been considered (or could at least be interpreted as such by taking the context of the study into account) as indicators for the achievement of one or more of these sub-goals?

In the following, the sub-goals (that we have already introduced above) are presented along with the process characteristics we have assigned to them.

Developing a strong understanding of the statement to be proved/ estimation of the truth:

In order to get familiar with the statement to be proved it may be useful to spend some time with the task (Zazkis et al., 2015), to read the statement repeatedly, and to reformulate it (Furinghetti & Morselli, 2009). Reformulation may include “using gestures, words, pictures, symbols, sketches, examples, and so on” (p. 3). In particular, it is assumed that examples provide insight into why a conjecture must be true or false (e.g., Ellis et al., 2017; Ozgur et al., 2017; Weber & Alcock, 2004). Process characteristics such as varying examples systematically, choosing examples with specific properties, and attempting to capture a broad range of examples may help to test the domain for which the conjecture holds true and to explore its limitations (Ellis et al., 2017). Weber and Alcock (2004) as well as Ellis et al. (2013) emphasized that examples have to be purposefully chosen in such a way that they only reflect properties that are consistent with the reference theory. The study by Antonini (2011) has shown that experts frequently only observe whether an example has the requested properties or not by applying the trial-and-error strategy. Collecting all ideas, previous knowledge, and examples that seemed useful and connected to the problem (Kidron & Dreyfus, 2014), as well as attempting to combine existing information into novel ways (Zandieh et al., 2014) are further process characteristics that may be helpful in becoming acquainted with the statement and its related concepts.

Inventing and formulating new conjectures or refining existing conjectures:

Systematically choosing a set of examples by varying one or more elements can be a first step to identify underlying structures and patterns from which a (new) conjecture may be inferred (Ellis et al., 2017; Pedemonte & Buchbinder, 2011). In particular, abductive inferences can be based on a set of examples (which then serve as facts) with the aim to construct a conjecture (Pedemonte & Reid, 2011). The use of algebraic expressions becomes relevant when trying to generate a generic example, which in turn can promote the formulation of a recursive rule (Pedemonte & Buchbinder, 2011). Referencing to Lakatos (1976), Komatsu (2011) claimed that a counterexample that discards an initial conjecture could be used to formulate a new conjecture by adding "a suitable lemma that will be refuted by the counterexample" (p. 149) as a condition to the discarded conjecture. In the study by Komatsu et al. (2014), it is also emphasized that the identification of counterexamples may be the starting point to invent new conjectures (that hold true for the counterexample). Boundary examples that represent extreme or special cases (e.g., Ellis et al., 2017) and thus target the boundaries of a conjecture can either be helpful for formulating a new conjecture (Ellis et al., 2013; Ellis et al., 2017) or for refining an existing one (Komatsu, 2011). According to Grenier (2013), the trial-and-error strategy can also be applied in order construct a conjecture, which is based on exploration by studying specific cases that are not obviously true.

Finding an adequate representation for the proof and an adequate proving strategy:

Within a socially-embedded argumentation, adequate representations for a proof may be found by listening to each other, correcting each other, and together negotiating them and the arguments that back them up (Mueller et al., 2012). Furinghetti and Morselli (2009) claimed that a learner has to switch flexibly between different representations in order to be able to distinguish between 'representations with future' and 'representations without future', that is identifying those that may be helpful for the future conjecturing and proving process. Yet, also staying within one representation system by, for instance, transforming algebraic expressions can be a promising strategy to "make explicit something that was implicit in the initial algebraic expression" (Martinez et al., 2011, p. 36). Regarding the aim of finding an adequate proving strategy, the process of switching forth and back may encourage to think about which direction might be easier to prove (Zandieh et al., 2014).

Examples-based justifications:

Different types of examples and counterexamples may be used to support one's own justification (e.g., Ellis et al., 2017; Komatsu, 2011; Lee, 2016). Refuting a conjecture by presenting a counterexample is a viable and complete proof (Ellis et al., 2017), other examples can only justify mathematical propositions that specify a finite set of objects. Lee (2016) explained that when "there are finitely many objects, the proposition may be proven true by verifying that each object satisfies the proposition, that is, each object is an example" (p. 28). He also pointed out that attempts to justify propositions that specify an infinite set of objects by verifying some examples leads to inductive reasoning. Reviewing a set of examples in order to find similar structures across this set of cases and building formality by replacing the numbers used in the examples with variables may help to construct a general argument (out of the set of examples) (Ellis et al., 2017). Several studies pointed to this type of process characteristic (i.e., process pattern generalization) and emphasized that one has to focus on the regularity in the process rather than on the regularity in the results (e.g., Ozgur et al., 2017; Pedemonte & Buchbinder, 2011).

Drawing inferences:

In semantic proof production, instantiations of mathematical objects that are meaningful to the prover may guide his or her formal inferences. The term ‘meaningful’ excludes cases in which an individual represents a mathematical concept by rewriting its definition without attaching meaning to it (Weber & Alcock, 2004). Antonini (2011) observed that the “transformations of signs (transformational process) seem[ed] to be guided by a concept image that allows a fruitful anticipation of some aspects of a final object” (p. 216). Collecting and combining various ideas, concepts, and examples may allow the development of such meaningful instantiations and concept images (cf. Herlina & Batusangkar, 2015; Kidron & Dreyfus, 2014), as well as the identification of appropriate warrants that are needed to support the argument one has constructed (Dawkins, 2015, p. 68). In addition to generating arguments that are based on meaningful instantiations, Weber (2004) claimed that students’ successful proof attempts could also be achieved simply by imitating the teacher’s actions or by applying a series of steps that have previously provided valid proofs. Proofs that are based on the application of procedures are termed procedural proof products. In syntactic proof production, “pushing symbols” (Furinghetti et al., 2013, p. 108) and “manipulating correctly stated definitions and other relevant facts in a logically permissible way” (Weber & Alcock, 2004, p. 210) represent the relevant process characteristics.

Structuring and organizing inferences:

It was assumed that studying the logical structure of the presented statement (e.g., Dawkins, 2012; Selden, A. et al., 2010), differentiating explicitly between the givens and the conclusion (Küchemann & Hoyles, 2006; Lee, 2016), and expressing them in an accurate symbolic language (Lin, F.-L. et al., 2004) are process characteristics that can encourage learners to generate a structurally sound chain of inferences. Writing “the formal-rhetorical part of a proof, that is, the part of a proof that depends only on unpacking and using the logical structure of the statement of a theorem and associated definitions” (Selden, McKee, & Selden, 2010, p. 2) may allow the prover to see where to start and end the body of proof. Systematizing already accepted statements and using existing relationships (Küchemann & Hoyles, 2006) as well as working forward from the givens and working backward from what one is required to show (Carroll, 1977; Küchemann & Hoyles, 2006) are process characteristics that were considered to be helpful for structuring one’s own proof. The step of establishing links between different arguments to generate a deductive chain may be guided by drawing analogies to familiar tasks (Dawkins, 2015). Controlling the chain of individual deductions can help to identify gaps in the overall proof (Furinghetti & Morselli, 2009).

Translating less formal to formal arguments:

The “attempt to evaluate what kind of relationship exists between the arrangement of links between and within inferences in an informal argument and the arrangement of links between and within inferences in a formal proof” (Zazkis & Villanueva, 2016, p. 329) may facilitate the achievement of structural continuity between less formal and formal arguments. Structural continuity can also be reached by linking the numbers used in the informal argumentation to the meaning of variables used in the deductive proof (Pedemonte, 2008) or by replacing unscientific or informal concepts with scientific ones (Dawkins, 2015). Weber and Alcock (2004) emphasized that it is important that informal representations only reflect the properties that are consistent with the reference theory when attempting to generate formal arguments out of them.

Communicating and presenting arguments:

“Formal language is precise, rigourous, and non-ambiguous” (Pastre, 1989, p. 273) and therefore often used to present arguments in an accurate way (Weber, 2004; Zaslavsky et al., 2012) to communicate with ‘qualified’ people (Pastre, 1989). Bagchi and Wells (1998) pointed out that logical and formal symbols are needed instead of words to ensure clarity. However, learners may draw sketches to demonstrate their understanding of the concepts involved in the proof problem (Plaxco & Wawro, 2015), generate examples to illustrate their claims that a conjecture is true (Ellis et al., 2017), and use non-formal language to express ideas, concepts, and methods (Pastre, 1989). Furthermore, Ellis et al. (2017) observed that some students use examples as illustrations to impart how a formal representation or a graph works, or what it represents in relation to the conjecture that has to be proved.

Resolving fixations/ avoiding errors:

When reaching an impasse, it may be useful to go back to previous proving steps, to revise the applied strategy, to check the chain of arguments, and to refine the language and the exterior form of the proof (Furinghetti & Morselli, 2009). Different researchers pointed out that reflecting on everything done so far and on one’s own strategy use may help to identify deadlocks (Huda, 2016; Savic, 2015b; Selden, A. et al., 2010). Provers that got stuck should try to remove incorrect calculations (Huda, 2016) and to make sense of their incorrect proof attempts (Savic, 2015b). Taking a walk, going to lunch in attempting to have successful insights (what is considered as a period of incubation), or sleeping on it are process characteristics that experts often use to recover from proving impasses. It may also helpful to move on in the lecture notes or to do other (mathematics-related) projects (Savic, 2015a). Other domain-general process characteristics that may encourage to overcome impasses are trying to remember how the problem before was solved in order to apply those proving techniques

(Huda, 2016) or transforming the ideas from another fields (Savic, 2015a) as well as splitting the theorems that have to be proved into two or more simpler ones (Pastre, 1989). Attempting to create a counterexample has also been seen as a promising process characteristic if one does not know how to proceed (Savic, 2015b). Anticipatory thinking (Huda, 2016),asking oneself certain monitoring questions (Savic, 2015a), and writing the formal-rhetorical part of a proof (Selden, A. et al., 2010) are process characteristics that have been assumed to be helpful to avoid errors. It may be necessary to correct a peer or to assist him or her in making sense of an argument that was originally expressed in an obscure or incorrect way when working collaboratively on a proof-related task and when mistakes occur (Mueller et al., 2012).

Generating a shared understanding:

Developing a shared understanding and conception of a proof or claim may result from parallel and successive internalization and externalization of ideas by individuals working together in a social context. Internalizing and externalizing processes are assumed “to promote changes in and refinements of both individual and shared mathematical notions” (Vidakovic & Martin, 2004, p. 490). Blanton and Stylianou (2014) claimed that transactive reasoning as a specific form of interaction (including explaining, critiquing, clarifying, requesting, and evaluating the arguments of others) fosters students’ capacity to generate arguments about complex, mathematical ideas, and, as such, has a positive impact on their learning of proof. In the study by Mueller et al. (2012), it has been shown that different types of collaboration (co-construction, integration, modification) may contribute to the exchange of ideas and to the creation of mathematical knowledge. These types of collaboration incorporate process characteristics such as externalizing one's own arguments as well as evaluating and integrating the arguments of others.

Producing understanding of the proof:

Some of the process characteristics attended to the sub-goal *generating a shared understanding* (such as explaining and evaluating arguments) may be listed here, again. However, we identified further process characteristics that were assumed to promote understanding and that have not been discussed from the proving as a social discursive perspective. For instance, generating examples or drawing diagrams may help to understand how a proof (or representation of a proof) works, and to understand its limitations (Ellis et al., 2017; Weber, 2004). In particular, examples with specific properties or boundary cases may promote understanding of the causal mechanisms behind a conjecture (Ellis et al., 2017). It is suggested to first attempt to understand the meaning of concepts in a proving process in order to be able to understand the whole proof (Uğurel et al., 2016). Trying to explain to oneself why the proof works (Weber, 2004) and assessing an argument against some criteria (e.g., is it convincing? does it provide evidence?) may help to understand the meaning of an argument (Mejía-Ramos & Inglis, 2009).

Table 6: *Sub-goals versus process characteristics. Lightly-shaded cells present the assumed or reported relations between the success within one sub-goal and the occurrence of specific process characteristics during the attempt to achieve the sub-goal.*

		Sub-goals										
		developing a strong under standing of the statement to be proved/ estimation of the truth:	inventing and formulating new conjectures or refining existing conjectures	finding an adequate representation for the proof and an adequate proving strategy	example-based justifications	drawing inferences	structuring and organizing inferences	translating less formal to formal arguments	communicating and presenting arguments	resolving fixations/ avoiding errors	generating a shared understanding	producing understanding of the proof
Process characteristics	example use	■	■		■	■			■	■		■
	working with informal representations					■			■			■
	using formal symbols and algebraic representations		■	■		■		■	■			
	reformulating	■		■								
	using abductive and inductive inferences		■									

	unpacking mathematical statements											
	organizing one's knowledge and existing definitions, concepts and structures											
	meta-cognitive processes											
	applying domain-general problem-solving strategies											
	transactive activities											

5.8 Discussion

5.8.1 Summary

Based on the fact that research on conjecturing and proving is growing rapidly, reflecting the importance of conjecturing and proving in mathematics education and the relevance of knowing the processes that are assumed or reported to be crucial for the success (cf. Mariotti, 2006; Stylianides, G. J. et al., 2017), our review had two main objectives: Firstly, we systematized the literature on conjecturing and proving using a topic modelling method. The algorithm returned 17 topics and a list of words composing those topics. By examining the articles and research reports that represent these topics, we have discerned that the research clustered to one topic often shares a common perspective on conjecturing and proving. Regarding this observation, we have been able to replicate the three perspectives on proving outlined by Stylianides, G. J. et al. (2017), namely the proving as problem-solving, proving as convincing, and proving as a socially-embedded activity, within the literature on conjecturing and proving. Yet, we have identified a new perspective, the discovery perspective on conjecturing and proving, within the literature and proposed a more fine-grained categorization of the perspectives introduced by Stylianides, G. J. et al. (2017). The discovery perspective emphasizes that conjecturing and proving can be used as a means for exploring, discovering, and inventing new mathematical results (cf. Villiers, 1999), especially in the field of geometry. Our analysis has shown, that the problem-solving perspective can be sub-divided into smaller categories such as the problem-solving perspective with a specific focus on strategy use, on affective and cognitive resources, or on approaches to overcome proving impasses. Studies that share the convincing perspective appear to be embedded in a specific educational context, either in the university or in the school context. Furthermore, our topic model indicates that the proving as a socially-embedded activity perspective is less present in the literature on conjecturing and proving. The articles and research reports of our document collection consist only of a small proportion of words from the "social/ collective argumentation" topic that typifies the social perspective on conjecturing and proving. This result is in line with Balacheff's (1988) critique that there is a too strong emphasize on the logical side of proof, disregarding its social importance as a means for communication. Stylianides, G. J. et al. (2017) also remarked that the proving as a socially-embedded activity perspective is less developed than the problem-solving and convincing perspective. Even though, we were open to consider further perspectives on conjecturing and proving, we have to confirm the observation of Stylianides, G. J. et al. (2017) that some articles and research reports do not fully fit within one of the three (in our case four) perspectives. However, our topic modeling approach has allowed us to systematize the literature on conjecturing and proving and to draw conclusions about the presence of specific topics that have been discussed in the literature in the context of conjecturing and proving. We have summarized the methodological orientations of the underlying studies of the most representative articles and research reports for each topic. This summary indicates that existing studies provide a rich qualitative basis on conjecturing and proving, but that quantitative findings resulting from the observations of larger populations are rare.

Secondly, we have analysed the most representative articles and research reports for each topic with regard to their claims and empirical findings about promising conjecturing and proving processes. We noticed that in the literature different types of processes related to conjecturing and proving have been presented and that the ways in which they have been described varies. Terms such as “exploration” (e.g., Komatsu et al., 2014; Mejía-Ramos & Inglis, 2009; Ozgur et al., 2017), “refinement of conjecture” (e.g., Komatsu, 2011, 2016), “producing understanding” (e.g., Ellis et al., 2017; Furinghetti & Morselli, 2009; Zazkis et al., 2015), or “justification” (Mejía-Ramos & Inglis, 2009; Zaslavsky et al., 2012) that are partially unprecise and difficult to operationalize have been used in numerous studies to delineate the intermediate steps that are needed to generate conjectures and to construct proofs. In some studies, they have been described rather as latent constructs than as directly observable processes (e.g., Küchemann & Hoyles, 2006; Mejía-Ramos & Inglis, 2009). However, we regarded these processes as necessary intermediate steps, as sub-goals within conjecturing and proving processes, which themselves require further, more fine-grained processes and which may be achieved in different ways. In total, we have been able to infer eleven sub-goals from the literature on conjecturing and proving. Some of these sub-goals, such as *developing a strong understanding of the statement to be proved/ estimation of the truth* or *finding an adequate representation for the proof and an adequate proving strategy*, are comparable to the four phases outlined in Polya (1945). Others, such as *inventing and formulating new conjectures or refining existing conjectures, structuring and organizing inferences,* or *communicating and presenting arguments*, occur in the phase model of Boero (1999) in a similar way. Moreover, we identified new sub-goals within the literature on conjecturing and proving such as *resolving fixations/ avoiding errors* or *generating a shared understanding*, which are of particular importance in specific situations (e.g., when an impasse is reached) or contexts (e.g., when an argumentation is embedded in a social context). As we were interested in finding out which processes have been assumed or reported to be helpful in achieving one or more of these sub-goals, we have expanded our analysis. We searched for process characteristics of conjecturing and proving that have been considered (or may be interpreted as such) as indicators of how to successfully accomplish these sub-goals. Our search resulted in a broad set of process characteristics that reflect multiple different ways in which successful conjecturing and proving processes may be carried out and that have been perceived as relevant from different perspectives on conjecturing and proving. Based on these results, we proposed a framework that takes both sub-goals and process characteristics of conjecturing and proving into account. By studying the literature with regard to the relationships between sub-goals and process characteristics, we found that the processes used to complete a specific sub-goal may vary widely. For instance, on one extreme, doing tasks unrelated to mathematics, taking a break, or going for lunch are processes that are assumed to be helpful in recovering from fixations (Savic, 2015a). At the other extreme, overcoming impasses may involve attempting to construct a counterexample, reflecting upon various proving techniques (Savic, 2015b), or testing the limitations of a conjecture (Ellis et al., 2013). However, we have derived from the literature that anticipatory and structural thinking (including the existence of a goal in mind when choosing and employing a specific process) are the two key aspects when attempting to achieve a sub-goal, as they can facilitate the achievement of a sub-goal or, if they lack, may hinder its achievement.

5.8.2 Limitations

The present review has some limitations that should be noted to enable an appropriate interpretation and use of our results. Firstly, some constraints stem from the methods we applied to create this review. We selected the literature by reading the headings and abstracts with respect to our inclusion criteria. Since this procedure is prone to error, we cannot be sure that we have included all the relevant literature on conjecturing and proving. Furthermore, this selection also leads to the fact, that some articles were included that relate more to machine learning than to mathematics educational topics. A topic modeling algorithm has been implemented to identify the major strands of research within the literature on conjecturing and proving. In general, the algorithm works by analysing the occurrence and combination of words within each paper (cf. Inglis & Foster, 2018). This method has turned out to be effective in providing an overview about the main topics that reflect the research on conjecturing and proving. However, the presence of particular words associated with conjecturing and proving is not sufficient to detect and summarize common claims and empirical findings about promising conjecturing and proving processes. Consequently, an in-depth analysis has been conducted in order to bring the claims and findings together. For this purpose and for each topic, we have qualitatively analysed the most representative articles and research reports. As we would not have been able to handle the deluge of data that would have resulted if we had tried to locate, read, and interpret all the relevant literature on conjecturing and proving, we confined the in-depth analysis to a total of 45 articles. Another limitation concerns the fact that we have not checked the consistency of interpretations or possible coding errors by applying a joint or double-coding procedure. In particular, this should not be disregarded, as the allocation of the "free codes" to the categories of sub-goals and process characteristics depends on the judgement and insights of the authors.

A second group of limitations has its origin in the literature we used. Only a minority of the reported studies used quantitative research methods or randomised experimental designs. Most results were inferred from data based on small sample sizes and qualitative research methods. Some may oppose the comparability and synthesis of qualitative research on the grounds that the results of single studies have been decontextualized and that the process characteristics of conjecturing and proving identified in one study are not directly applicable to other studies or research contexts (cf. Thomas, J. & Harden, 2008). In addition, we observed that the articles and research reports that represent a topic were often been written by authors who belong to the same research group. This might be one reason why they share similar perspectives on conjecturing and proving, similar theories, beliefs, methods, and data sets (cf. Lakatos, 1978, as cited in Inglis and Foster (2018). Finally, as stated in the beginning of this review, it can be critically remarked that we did not distinguish between the process characteristics that were inferred from the observation of expert mathematicians and those that were inferred from data of undergraduate students (cf. Reif, 2008).

5.8.3 Implications for research and teaching

We close this paper by proposing some implications and recommendations for future research and teaching based on our results.

Even though conjecturing and proving processes often take place in mathematics classrooms (e.g., Vidakovic & Martin, 2004; Yackel & Cobb, 1996) or in form of mathematical debates (Alibert & Thomas, 1991), the results of our topic modeling approach indicate that studies that conceptualize conjecturing and proving as socially-embedded activities are rare. Stylianides, G. J. et al. (2017) critically remarked that this perspective on conjecturing and proving is still poorly evolved and not yet coherent. Research on how individual proving ideas develop in a social context appear to be underrepresented. We claim that research in this area has to be extended and that an alternative perspective that takes individual and social characteristics of conjecturing and proving into account should be adopted.

Most of the studies described in this review have used exploratory research approaches (such as grounded theory) to understand the processes that are needed to successfully construct conjectures or generate proofs. These studies provide a fruitful qualitative basis for future research directions and for the formulation of hypotheses about promising conjecturing and proving processes. For instance, based on the findings of the study by Ellis et al. (2017) and of the study by Weber and Alcock (2004), it can be hypothesized that students who choose and use examples purposefully (in the way that they systematically vary examples, search for similar mathematical structures, and build formal representations out of the examples) are more likely to be successful in producing semantic proofs than students who only pick some random examples. However, confirmatory research about conjecturing and proving processes is still missing. We claim that a quantitative validation of the results found in the qualitative studies would facilitate a more generalizable picture of conjecturing and proving processes.

Furthermore, the findings of our in-depth analysis support the observation by Dawkins and Karunakaran (2016) that there exists a large number of studies on proof-orientated behaviour that make content- and context-independent claims about promising conjecturing and proving processes. Discussing conjecturing and proving processes in a content- and context-independent way can be criticised on the grounds that the framing of research questions and the methodological choice of collecting and analysing data may non-trivially influence the nature of the phenomena observed and thus the studies' findings. For instance, we assume that students or mathematicians who receive support (cf. Komatsu, 2011), who have no time restrictions (cf. Herlina & Batusangkar, 2015; Savic, 2015a), who work on geometry tasks (cf. Küchemann & Hoyles, 2006), or who are allowed to use lecture notes (cf. Selden, A. et al., 2010) may behave differently than those who get no assistance, who work with time-limits (cf. Lin, F.-L. et al., 2004), or who have to solve analysis tasks (cf. Kidron & Dreyfus, 2014).We intend to sensitize the research community on the role that particular measurement methods or other context factors may have on the studies' results about conjecturing and proving processes.

By distinguishing separate but related categories of sub-goals and process characteristics of conjecturing and proving, our framework not only highlights the complexities associated with conjecture generation and proof construction, but also offers a way to understand how and why the occurrence of specific process characteristics may increase the probability of being successful. The sub-goals describe the intermediate steps that are considered to be necessary for generating interesting conjectures and constructing valid proofs. The process characteristics are observable and assumed to be potential indictors for the success within a certain intermediate step. As the intermediate steps such as exploring the problem situation or producing understanding of the proof have been listed as (latent) sub-goals in several studies, mathematics educators and researchers will need to operationalize these constructs. The process characteristics we identified may be used to operationalize the sub-goals within conjecturing and proving, and therefore may be valuable for analysing and assessing students' conjecturing and proving processes. Furthermore, the proposed framework may be adapted for teaching purposes. We suggest that teachers and lecturers should introduce the process characteristics in combination with the associated intended sub-goals that might be accomplished by employing them. Pointing to the different types of process characteristics can give students insights into how the sub-goals considered necessary for success may be the consequences of the appearance of certain process characteristics during the conjecturing and proving processes. Even though it is hard to determine the relative importance of each of these sub-goals respectively process characteristics of conjecturing and proving, as the types of processes successful provers engage in may vary according to different contexts, we claim that our framework may provide guidance for enhancing the learning and teaching of conjecturing and proving in (undergraduate) mathematics classes and for systematically analysing students' proving behaviour.

6 A rating scheme for assessing process characteristics of collaborative conjecturing and proving

The summary and findings of our literature review show a comparatively detailed picture of potential relationships between (collaborative) conjecturing and proving processes and proof performance. We observe a large body of qualitative research offering diverse hypotheses about relevant sub-processes and process characteristics of conjecturing and proving, but little systematically generated evidence about the importance of the hypothesized sub-processes and process characteristics. Therefore, we see the need for empirically investigating the impact of a set of process characteristics of collaborative conjecturing and proving on the quality of the resulting product.

One of the central goals of this dissertation was develop an instrument to describe and analyse collaborative conjecturing and proving processes along several theory-based process characteristics inferred from the mathematics educational as well as the psychological and the Learning Sciences literature. A high inference coding scheme was designed, based on existing guidelines and an extensive literature search. But beforehand, a theoretical excurse is presented to illustrate the considerations that guided the development of this coding scheme.

This chapter describes the decisions that had been made before the data collection and data coding was realised. We demonstrate how process characteristics were chosen and operationalized, present the whole rating scheme, and provide an overview about the rater-training. We complete this chapter by illustrating how the rating scheme was applied in our empirical studies and which further instruments have been used to assess undergraduate students' mathematical argumentation skills.

6.1 Why using "real-time" recordings and high inference coding strategies to analyse collaborative conjecturing and proving processes?

To investigate prospective undergraduate mathematics students' conjecturing and proving processes, we decided to use computer-supported learning environments that allowed to record students' verbal face-to-face interactions and their written utterances. All screen and audio activities were recorded by the laptops and transformed into a video file. In general, we assume that the use of "real-time" recordings of students' interactions during collaborative proof construction activities provides several advantages over other methodologies (such as questionnaires, interviews or real-time observations, etc.) (cf. Roth, 2009):

- Video-recordings as a "real-time" data collection technique allow comprehensive analyses of collaborative conjecturing and proving processes

We assume that video-recordings that capture "real-time" conjecturing and proving processes allow the identification of the process characteristics that may lead to an impasse as well as the process characteristics that may lead to success. "Real-time" conjecturing and proving data are also expected to provide information regarding what is going on during collaborative conjecturing and proving processes, how impasses occur, and which activities are promising to overcome them (cf. Savic, 2015a). Moreover, this kind of data may allow to describe the process characteristics that positively influence the generation of conjectures and construction of proofs and the characteristics that predict different patterns of peer collaboration.

- Video-recordings of screen and audio activities allow precise and subtle analyses of collaborative conjecturing and proving processes

As collaborative conjecturing and proving processes are very complex and involve several activities, such as generating examples, applying definitions, or selecting and evaluating the arguments of others, researchers can reach their limits to perceive all aspects that occur simultaneously. In order to avoid missing important details and information that may explain students' success or troubles in formulating conjectures and generating proofs, video-recordings can be stopped, slowed down and broken down in subsections. Consequently, researchers can code students' collaborative conjecturing and proving processes in multiple passes, and review every interaction the students have made at several times (cf. Roth, 2009).

- Video-recordings of screen and audio activities allow to achieve high inter-rater agreements and the objective coding of collaborative conjecturing and proving processes

Achieving a high inter-rater agreement is a major challenge in the context of assessing the quality of students' collaborative conjecturing and proving processes. In general, quality judgments require inferences based on the observable data that go beyond counting directly observable events or assigning characteristics to a particular category (cf. Seidel, 2005). A set of video-recordings of previous studies can be used for rater-trainings. Pointing to specific instances occurring in one of the video-recordings can help to establish coding rules and serve as anchor examples. Disagreements can be resolved by explaining one's coding decision by reviewing the video-recording together. Moreover, a precise coding procedure (see aspects listed under 2) can help increase the likelihood of high levels of inter-rater agreements (cf. Roth, 2009).

- Video-recordings of screen and audio activities allow both qualitative and quantitative analyses of collaborative conjecturing and proving processes

Video-recordings of students' collaborative conjecturing and proving processes can be qualitatively analysed (e.g., do students' accurately present their arguments by providing adequate warrants, do they equally contribute to the collaboration process by exchanging and evaluating each other's idea, etc.). These video recordings can be used to capture multiple qualitative descriptions of collaborative conjecturing and proving processes, but they also allow for more quantitative analysis (e.g., how often do students present new ideas or generate examples, what is the average length of students' utterances in the collaboration process, how many times do students formulate questions, etc.). Another advantage is that video-recordings receive data that can be re-analysed and used for a variety of purposes in the future. For instance, researchers may select some of the video-recordings to provide additional evidence to communicate their results within the mathematics educational community, or to use them as best-practice examples for teaching purposes (cf. Roth, 2009).

Video-recordings initially constitute raw data material. A further decision regarding the coding strategy has to be made. As our research questions address the assessment of the quality of collaborative conjecturing and proving processes, the use of high inference rating scales is reasonable (cf. Clausen, Reusser, & Klieme, 2003; Seidel, 2005). Low inference coding techniques are more suitable for counting how often an event or specific characteristic occurs (e.g., checklists) or for classifying a specific characteristic into one category (e.g., category systems). In general, their coding instructions can be formulated very clearly and achieving a high reliability is much easier than with high inference coding strategies. Due to the fact that instruments such as checklists or category systems often split observational events or characteristics down to the smallest detail, a global view on the underlying construct that should actually be measured is often lost. Based on the experience of previous studies that focussed on process-product correlations within the educational science, the use of high inference coding strategies appeared to be preferable. For instance, the IPN video study has also contributed to strengthen video-recordings as a methodological design and to apply high inference coding techniques (cf. Seidel et al., 2005). The use of global ratings as a particular high inference coding technique enables to capture the content and structures of the underlying construct in a valid way (cf. Clausen et al., 2003; Gartmeier et al., 2015; Newble, 2004). Rating scales produce data that can be handled as approximately interval-level, especially if the endpoints of the scales are considered as the extremes of a continuum (Wirtz & Caspar, 2002). Moreover, the study of Meier et al. (2007) also showed that rating scales provide an adequate technique "to evaluate the quality of collaboration processes on a relatively global level" (p. 71) and that their application is time-efficient, since the transcription of students' dialogue is not necessarily required. From this perspective and with regard to our research questions and aims, we think that developing high inference rating scales that take the entire collaboration process into account appears to be an adequate coding strategy for analysing video-recordings of prospective undergraduate mathematics students' collaborative conjecturing and proving processes.

6.2 Developing of a high inference rating scheme

This section will go into detail about the development procedure of the rating scheme we created to describe and analyse (prospective undergraduate) students' collaborative conjecturing and proving processes. We followed the guidelines proposed by Seidel and colleagues (2005): In a first step, a theoretical foundation for the high inference rating scales was generated. We conceptualized collaborative conjecturing and proving processes as a two-dimensional construct comprising the components: (i) individual-mathematical process characteristics and (ii) social-discursive process characteristics. Since we were interested in both components of collaborative conjecturing and proving processes, our literature search included research from mathematics education as well as from psychology and the Learning Sciences. We aimed to identify a set of theory-based process characteristics of collaborative conjecturing and proving, predicting the success of collaborative conjecturing and proving processes. Process characteristics should be defined with regard to our main assumption, based on the hypotheses put forward in the literature, that students engaged with a high level of quality in one or more of these process characteristics are likely to formulate an interesting conjecture and a valid proof (valid in the sense of being accepted by the corresponding mathematical community).

Our literature search consisted of two parts: a structured search using specific key words in databases (such as the keyword “successful” AND “proving”) and a slightly more unstructured literature search performed by cross-referencing the articles that were found to be interesting within the structured search. The structured search was a part of the literature search for our research review as described in Chapter 5 (Study I). All potentially relevant literature was selected and we deduced seven process characteristics of collaborative conjecturing and proving processes, three related to the individual and four related to the social-discursive component of mathematical argumentation. Table 11 describes the process characteristics we inferred from the literature. Subsequently, the unity of analysis was chosen. Initially, we decided to make qualitative judgements over the entire collaboration process for each of the seven process characteristics. The formulation of the rating scales resulted from our identified set of process characteristics and our theoretical framework of mathematical argumentation skills (e.g., including sub-processes of exploration and systematization; cf. Chapter 2). The scaling of the individual rating scales (five quality levels) was determined in accordance with the methods used in other studies (cf. Gartmeier et al., 2015; Seidel, 2005) and in coordination with the video recordings at hand. In order to achieve a high interrater agreement, we created detail coding rules as well as descriptions and anchor examples for each rating scale (Langer & Schulz von Thun, 2007; Seidel, 2005). The levels of the rating scales roughly correspond to German school grades (1 = excellent to 5 = fail). Video-recordings that were used for developing and testing were not included in the final data analyses (Seidel, 2005). The rating scales were tested and validated with video-recordings from previous studies by comparing the judgments of two to four researchers (that discussed the rating scales in a group several times). Based on the observations during the test phase, it was decided to change the unit of analysis by dividing the entire collaboration process into two parts to reduce the complexity of the coding process. After the rating scales were developed, the rater training started and inter-rater agreement was checked using a randomly selected sample of 10 video-recordings. A more detailed description of the rater-training is presented in the next section. The training was finished after both raters agreed that their codes are based on a shared theoretical understanding and the inter-rater agreement appeared to be good (ICC ≥ 0.6). As a last step, all 49 video-recordings of prospective undergraduate students' collaborative conjecturing and proving processes were analysed. Figure 8 shows the entire development procedure for the rating scheme.

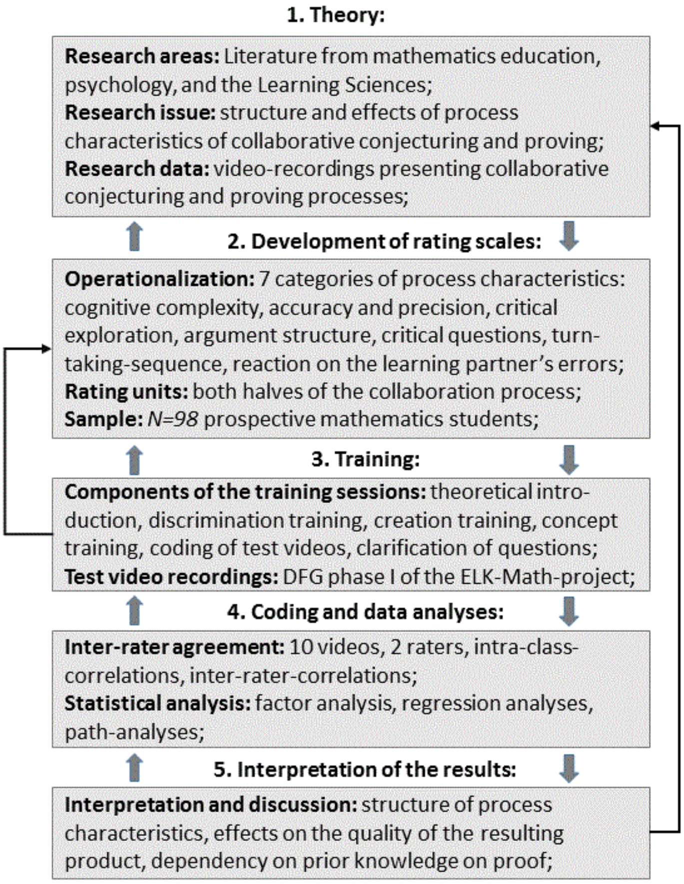

Figure 8: Development process for rating scales (adapted from Seidel, 2005)

In the following, we provide an overview of the fundamental coding rules and the initially constructed rating scales, and present one rating scale in more detail. It has to be noted that the descriptions of the rating scales and of the coding rules are not to be seen as a finalized product, since they only acted as support for decision making for the coding and they are to a certain degree specific to our concrete context (undergraduate student population, task, collaborative setting). However, the descriptions of the rating scales and coding rules served as basis for the rater training sessions and were adapted during the training (cf. Seidel, 2005). The final version of the coding scheme can be found in the Appendix.

6.2.1 A brief description of the rating scheme

Instruction and basic rules:

The coding scheme is a guideline for the rating of undergraduate students' collaborative conjecturing and proving processes. It is structured according to our conceptualization of collaborative conjecturing and proving processes, consisting of two facets, one individual and one social-discursive component. The coding scheme comprises seven rating scales, four related to individual and three related to social-discursive process characteristics. We established the following rules:

- The rating-scales serve to assess students' collaborative conjecturing and proving processes along selected theory-based process characteristics.
- The individual rating scales, the concepts involved, and their quality levels are explained scale-wise.
- For each rating scale: The pass requirement for a quality level is that the student fulfils all its criteria. If this "minimum requirement" is not satisfied, a code for one of the lower levels has to be given.
- The raters watch the video-recording up to one half of the collaboration process and evaluate this part of the recording for the seven process characteristics, after that, the second part is watched and rated. The raters can stop and replay parts of the video-recordings as required.
- Every learner is individually assessed along the seven rating scales.
- If one half of the collaboration process contains few explicit content for a single learner and process characteristic, so that a specific rating is not possible, then there is no evaluation for those process characteristics. Code "missing data" (9) will be assigned to the coded segment for this characteristic.

6.2.2 Overview of the rating scales

The rating scales were developed in order to assess (i) to what extend ideas and arguments are formulated correctly and precisely, and to what extend arguments are reworked in the case of a presumed error or wrong direction (*accuracy and precision*); (ii) to what extent different ideas are developed, combined, and linked to the definitions and underlying concepts involved in the proving problem, and to what extent the learner brings in new perspectives on the collaborative conjecturing and proving process (*cognitive complexity*); (iii) to what extent the learner explores both what is available to use (without having any initial idea of how to proceed) and what could be proved, and to what extent the learner critically investigates different conjectures by generating examples, counter-examples, and testing the constraints (*critical exploration*); (iv) to what extent the learner formulates structurally sound arguments by basing claims on data and using warrants to justify the link between the data and the claim, and to what extent the learner defines the scope of the argument by using qualifiers (*argument* *structure*); (v) to what extent the student questions the learning partner's ideas, arguments and proving strategies to comprehend his/ her approach better, and the level of elaboration of these questions (*critical question*); (vi) to what extent the learner actively contributes to longer phases of a coherent joint discourse by exchanging ideas and taking the learning partner's contributions into account, and to what extent the learner refines arguments until a joint argument is built (*turn-taking*); (vii) to what extent the learner identifies errors and impasses within the learning partners' arguments, and to what extent these errors are explained and alternative solution steps are proposed (*reaction to the learning partners' errors*). A more detailed description about the process characteristic *argument structure* is presented below as an excerpt from the coding scheme (and translated to English).

Quality levels of the process characteristics "argument structure"

Question: To what extent does the learner formulate structurally sound arguments during the proving discourse? Does the learner base his or her claims on warrants that justify the link between the data and the claim? To what extent does the learner define the scope of an argument by using qualifiers?

Criteria: Arguments that are central within the collaboration process are taken into account. These arguments are evaluated with regard to their structural elements. A central criterion is that claims are based on data and warrants that give support for the link between the data and the claim. Depending on the context of the discourse, the argument has to be qualified in order to demonstrate the degree of certainty with which a conclusion is drawn.

In each case, it is not relevant whether the claims, warrants and data are correct, but whether they are structurally complete in themselves.

Moreover, it is not central that every argument is complete, but that the parts that are essential in the discourse and that are not already obvious elsewhere are explicated.

1	The learner's arguments are detailed in their structure throughout the collaboration process, in particular at all phases of the discourse where it is possible and helpful. Claims are explicitly connected to warrants, and qualifiers are adequately used with regard to the type of warrant. Warrants usually go beyond empirical or intuitive support, and are related to the reference theory.
2	The learner's arguments go beyond the formulation of a "blank" claim. Only in a few cases, arguments lack cues as to the data on which the claim is based and with what degree of certainty the conclusion is drawn. In addition to empirical or intuitive support, warrants are mainly related to the reference theory.

3	The learner's arguments mainly go beyond the formulation of a “blank” claim. However, at a few key points of the discourse, where it would be necessary, there is a lack of cues as to the data on which the claim is based and with what degree of certainty the conclusion is drawn. Warrants are mostly limited to empirical support.
4	The learner’s arguments that go beyond the formulation of a “blank” claim are rather rare. At many key points, there is a lack of cues as to the data on which the claim is based and with what degree of certainty the conclusion is drawn. Warrants are mostly limited to empirical support.
5	Even in the case of central arguments and phases of the discourse in which it would be appropriate and necessary, the learner puts forward claims that lack warrants and qualifiers. The arguments produced by the learner are mostly limited to the formulation of claim, possibly with few exceptions.

6.3 Training of the raters

The use of high inference rating schemes demands an extensive rater-training in order to develop a joint understanding on the conceptualizations of the rating scales and to increase the reliability of coding. Two raters participated in a three-day rater training, in three further meetings the agreement between the two raters was checked by using, in total, a sample of 10 “training video-recordings” that were not included in the final analyses. The three-day rater training consisted of six sessions. The structure and composition of the rater-training was designed according to the guidelines proposed by Wiesbeck (2015) as well as by Langer and Schulz von Thun (2007). Session I started with a theoretical introduction of the process characteristics in order to acquire knowledge about their conceptualizations. Raters were asked to solve the conjecturing and proving problem, which was presented to the students as task, on their own and to read the coding scheme. Subsequently, there was time for group discussion. Session II incorporates the “discrimination training”: Raters were requested to sort three video-recordings with regard to their understanding of “good” collaborative conjecturing and proving processes. This training component severed as starting point to become familiar with the video-recordings at hand and to develop an idea of how students deal with the task and interact with each other. In session III, the “creation training” was conducted: the mediocre video-recoding from the discrimination training of the second session had to be transformed into a good respectively bad version, for each rating scale. Afterwards, “good” and “bad” versions of the process characteristics of collaborative conjecturing and proving were discussed. Session IV incorporated a “concept training”: Raters watched one video-recording in group and listed all observable processes they regarded as important. These conjecturing and proving processes were jointly assigned to the rating scales and appropriate scores. The purpose of this session was to become acquainted with the coding scheme, to allocate students' observable conjecturing and proving processes to the rating scales, and to identify typical examples that may serve as anchors for further coding. In session V, the two raters coded one new video-recording along all process-characteristics. They documented all questions and difficulties that arose during the coding procedure. After the run, the ratings of the process characteristics as well as the questions and difficulties were discussed in group and alternations were made. Session VI started with repeating the coding rules, especially those that have been set up during the rater-training. Moreover, frequent rater errors as well as techniques how to avoid them were discussed. After this first training phase, a second phase followed by a trial run in which 5 video-recordings were rated (three video-recordings from collaborative working sessions that were not part of our empirical studies and two video-recordings from the session that constitute our database). For each video-recording that served as training material, the ratings of both rater were compared (without any statistical analyses). Cases where ratings diverged were discussed and the corresponding sequences of the video-recordings were re-analysed. The aim of this second phase of training was that the raters got more experienced with the coding scheme and the video-recordings at hand. Final questions were clarified and a further set of 10 video-recordings was selected for calculating the inter-rater agreement. These video-recordings pertained to our database. The inter-rater reliabilities (ICCs) and inter-rater correlations for all rating scales are presented in study II (Table 7). The inter-rater agreement across all seven rating scales (process characteristics) of the coding scheme was $M_{\text{ICC}} = .90$ ($SD_{\text{ICC}} = .11$). An overview of the entire rater training and its training components can be found in the Appendix.

6.4 Study design and instruments

We applied the newly developed rating scheme to process-data from a sample of 49 collaborating dyads. The collaborative working session in which data on students' conjecturing and proving processes were collected was embedded in a two-week voluntary preparatory course for prospective university mathematic students. This preparatory course incorporated 14 lectures and 14 tutorials on basic mathematical topics (e.g., propositional and predicate logic, proof techniques, number theory, functions, induction and recursion), four test sessions (including pre- and post-tests to assess the participants' knowledge on proof and their domain-general knowledge on argumentation) and 6 collaborative working sessions in total. All six collaborative working sessions and testing phases were also part of a large study, the *ELK Math Study - Effects of heuristic worked examples and collaboration scripts on the acquisition of individual and discoursive components of mathematical argumentation skill* (Fischer, Reiss, Ufer, & Kollar; funded by DFG 2011-2018).

In the tutorials, conversations were mainly based on students' work. Tutors demonstrated how they construct a proof, encouraged students to develop their own solutions, and hence, gave students the opportunity to experience different types of proofs. In the collaborative working sessions, students were paired in dyads and worked collaboratively in a computer-supported learning environment on different mathematical conjecturing and proving tasks, one task per session. The computer-supported learning environment has already been used in prior studies (cf. Kollar et al., 2014; Vogel et al., 2016). To avoid major discrepancies in the learning prerequisites of students working within one dyad, we decided to form homogeneous dyads with respect to their prior school achievements (cf. Webb, Nemer, & Zuniga, 2002).The first collaborative working session (that started on day six of the course and took around 90 min) represents our data basis for evaluating our rating scheme and for investigating our empirical research questions. During the session, students did not receive any feedback as we were interested in how students collaboratively generate conjectures and construct proofs without any instructional support. Nevertheless, in the tutorials, before the beginning of the collaborative working sessions, participants got the opportunity to formulate conjectures to a different mathematical topic, to construct proofs, and to receive feedback from tutors. In the collaborative work sessions, the learning partners of each dyad was seated on opposite sides of a table and were equipped with laptops and graphic tablets. The computer-supported learning environment allowed them to exchange written ideas and arguments by using a graphical chat, as well as to communicate to each other verbally face to face. To assess the final product (the formulated conjecture and generated proof) of students' collaborative conjecturing and proving processes, students were asked to write down an individual solution on a blank paper at the end of the collaborative working session. The final product was evaluated with regard to the correctness and creativeness of the formulated conjecture, the mathematical ideas that became visible in the solution, the soundness of the produced chain of arguments, and the correctness of the formal representations used to present the arguments.

In our empirical study (see Chapter 7 and Chapter 8), students worked on the following open-ended conjecturing and proving task: *"Take four consecutive numbers, multiply them, and add one. Repeat this and try to find similarities. Formulate a conjecture and prove it!"* The fact that students were requested to generate their own conjectures makes their work different from the typical proving tasks that are usually presented in the school or university context, where students often have to develop arguments to support a claim they might never have thought of before (cf. Douek, 2007). As it was a further research issue to examine how students' prior knowledge on proof affects their collaborative conjecturing and proving processes, we adapted the argumentation-skill-test designed by Reichersdorfer et al. (2012) and administered it on day four of the course during a testing session to obtain a measure of the participants' prior knowledge on proof in elementary number theory. This test was developed to capture four facets of mathematical argumentation skills: “technical proof skills”, “flexible proof skills” and “evaluations skill for true or false mathematical statements”. A more detailed description about the test and how it was coded, can be found in chapter 8 as well as in the Appendix.

7 Study II

Good collaborative conjecturing and proving processes – The structure of individual-mathematical and social-discursive process characteristics

7.1 Abstract

One of the central goals of university mathematics programs is improving students' skills to formulate mathematical conjectures and to provide evidence for the truth or falsity of these conjectures. Frequently, such argumentation processes occur in collaborative settings. Generating conjectures and proofs thus requires both individual-mathematical activities such as exploring the problem situation and developing deductive lines of arguments as well as social-discursive activities such as engaging in meaningful mathematical discussions. Even though both types of activities are intensively discussed in the literature and are often used to conceptualize the quality of collaborative conjecturing and proving processes, it is unclear to what extent they can be seen as independent from each other. Furthermore, there is still only limited knowledge regarding the question of how good mathematical conjecturing and proving processes can be systematically described. We introduce an analytic rating scheme that outlines collaborative conjecturing and proving processes along seven theory-based process characteristics that take an individual-mathematical and social-discursive perspective on argumentation into account. In a study with *N=98* incoming mathematics students, this new rating scheme was used to investigate the empirical relations between these characteristics of collaborative conjecturing and proving processes. Results showed that individual-mathematical and social-discursive process characteristics can be clearly distinguished empirically. We discuss ways how to operationalize different characteristics of collaborative conjecturing and proving processes and implications for further research.

7.2 Introduction

Conjecturing and proving are challenging for students at all educational levels (e.g., Yang, 2012). Even undergraduate students in mathematics often struggle to construct proofs and to communicate them correctly and with precision (Epp, 2003). Prior research has documented different types of errors and underlying misconceptions (e.g., Healy & Hoyles, 1998; Moore, 1994; Selden, A. & Selden, 1987; Selden, A. & Selden, 2008), but provides few directions in terms of which activities are good for the generation of valid conjectures and acceptable proofs. "Good" means that these activities are sufficient or even necessary to develop a connected sequence of true statements that is structurally valid and justifies the truth of the formulated conjecture (cf. Ozgur et al., 2017; Stylianides, A. J., 2007).

The exploration of conjectures and the construction of proofs are central practices within mathematics (e.g., Lakatos, 1976; Lin, F. L. et al., 2012). Therefore, it is essential that university students are able to implement conjecturing and proving processes successfully on their own as well as in collaborative settings (e.g., Lakatos, 1976; Lin, F. L. et al., 2012). Following Roschelle and Teasley (1995), we talk about *collaborative* when students or mathematicians working in small groups or dyads strive to develop and maintain a shared understanding of the (proving) problem.

In this contribution, we study undergraduate mathematics students' collaborative conjecturing and proving processes. We consider collaborative conjecturing and proving as a specific type of argumentation incorporating the formulation and exploration of a mathematical conjecture, the generation of adequate arguments for or against it, the combination of these arguments into a deductive proof, as well as the evaluation and integration of arguments produced by others (cf. Reichersdorfer et al., 2012). According to previous conceptualizations of mathematical argumentation (e.g., Kollar, et al., 2014), we distinguish between two components: One related to *individual-mathematical* activities such as deducing conjectures from examples (Koedinger, 1998), "unpacking the conclusion" (Selden, J. et al., 2014, p. 246), selecting and enhancing arguments (Boero, 1999), or developing different proof strategies (Zazkis et al., 2015). The other component focuses on *social-discursive* activities, which refer to more general, dialogical argumentation skills and stem from more a domain-overarching strand of research in psychology and the Learning Sciences (e.g., Asterhan & Schwarz, 2007; Kollar et al., 2007). The social-discursive component points to the facts that mathematical knowledge generation is often embedded in a social context (e.g. seminars, tutorial sessions) and that participating in mathematical discourse is vital for collaborative conjecturing and proving. From this perspective, activities such as exchanging ideas and feedback (Reinholz, 2016), eliciting knowledge from the learning partner (Weinberger & Fischer, 2006), building upon each other's arguments (e.g., Chi & Wylie, 2014; Teasley, 1997), and providing evidence for the weakness or incorrectness of a partner's argument (Asterhan & Schwarz, 2007) are important for a fruitful argumentative discourse with the goal to acquire knowledge, but mostly conceptualized independently from the specific mathematical content.

This study is concerned with the development and empirical testing of an instrument that makes the differentiation between the individual-mathematical and social-discursive component possible. (1) We will present a set of seven process characteristics of collaborative conjecturing and proving that are assumed to be relevant for successful mathematical argumentation from a theoretical point of view based on current literature. (2) Furthermore, we will provide data on the empirical structure of the extracted process characteristics for one exemplary conjecturing and proving task.

7.3 Research on collaborative conjecturing and proving processes

7.3.1 Individual strategies and activities during conjecturing and proving

Conjecturing can be considered as a creative work including different experimental activities such as investigating examples and counter-examples, discovering new logical relations between previously unrelated ideas and arguments as well as drafting and formulating conjectures (Yang, 2012). These activities seem to be crucial for a broad range of disciplines, but still they differ across domains. Dissimilarities include, for instance, what counts as a valid conjecture and as supporting evidence (Lin, F. L. et al., 2012). In contrast to other domains such as biology, medicine or politics, the ideal evidence in university mathematics is a chain of deductive arguments based on axioms and definitions (Fischer et al., 2014). Mathematicians agree that empirical, intuitive and authoritative arguments have strong limitations (Weber et al., 2014). Following Stylianides, G. J. et al. (2017), proving is the process of constructing a sequence of arguments for or against a mathematical conjecture that is characterized by using only previously accepted statements, theorems and definitions, valid forms of reasoning and adequate forms of notations. What can be regarded as 'valid' or 'adequate' is defined by the respective mathematical community and is partially dependent on the specific context of the proof construction process. The proof construction process itself has been described by different researchers in the form of models that are based on theoretical assumptions (e.g., Boero, 1999) or on self-reports (e.g., Schwarz et al., 2010).

Boero (1999) created a process model of proof consisting of various phases that start from exploring the problem situation to formulate a mathematical conjecture and end up in writing down a proof in a readable way that corresponds to the sociomathematical norms. This expert model refers to the assumption of cognitive unity (Garuti et al., 1998), which outlines strong relations between the activities of conjecturing and proving. Schwarz et al. (2010) suggested that three main activities related to proof construction – "enquiring", "proving" and "inscribing proof" – should be differentiated. The first activity concerns making sense of the problem, establishing conjectures and intermediate hypotheses (subgoals) for the proof (cf. Heinze, Cheng, Ufer, Lin, & Reiss, 2008a). Developing a deductive chain of theory-based arguments that connects the prerequisites with the claim of the conjecture represents the second activity. The third activity includes checking the logical integrity of the proof, and communicating it with formal precision.

These frameworks give some indications about which activities may occur during the proof construction process, and how an ideal proof construction process may look like. Since empirical-inductive and formal-deductive steps are incorporated in both models, we conclude that exploratory activities, checking the consistency between the mathematical concepts related to the proof problem and one's own arguments, as well as unpacking the logical structures of statements are crucial within conjecturing and proving processes.

Exploratory activities such as generating examples (e.g., Koedinger, 1998), reflecting on familiar problems, and associating similarities between them (Selden, A. & Selden, 2013a) can help students to understand the problem, find a conjecture or devise a plan for solving the task (e.g., Mills, 2014; Polya, 1945). Sandefur et al. (2013) the strategy of creating examples in university students' efforts to prove or refute a mathematical conjecture. In their study, students worked in small groups on number theoretic tasks. Although most groups tried to find examples that would provide some conceptual insight into the structure of a statement, example use varied among the groups within problems. The authors assumed that the use of example-based reasoning strategies depends on students' experience, their personal example space, and the way problems are presented. Drawing diagrams can be considered as a further exploratory activity that might help discovering new ideas or getting empirical evidence for the truthfulness of an argument (Gibson, 1998). For some mathematicians, the exploration of examples or visual arguments is an essential part in coming to understand new concepts and to justify new theorems, while others construct proofs that are entirely based on the manipulation of symbols within the representation system of the given problem (Alcock & Inglis, 2008). Certainly, it is an advantage to be able to use both strategies, but many students and also some professionals obviously prefer one type of reasoning (Zazkis et al., 2015).

The consistency between the mathematical concepts that are related to the given proof problem and one's own ideas constitutes a central aspect within the proof construction process (Mariotti, 2006). This means that students have to connect the formal definitions of the concepts to the instantiations they use for their argumentation. The mathematical objects that make up their arguments may only have properties that conform with the formal theory (Weber & Alcock, 2004). To achieve the required consistency, activities such as using different representations (Boero, 1999; Ufer et al., 2009) or operable definitions (Selden, J. et al., 2014) may be helpful. These types of activities may depend on students' individual conceptual understanding that, in turn, enables them to monitor their performance and to identify their own impasses (Ohlsson & Rees, 1991).

Selden, A. et al. (2010) claimed that the enactment of "behavioral schemas" (p. 205), which are partly procedural knowledge, affect the use of logical structures. Unpacking the logical structure of a statement can be considered as a first important step within the proof construction process, since "the logical structure of a mathematical statement is closely linked to the overall structure of its proof" (Selden, A. & Selden, 2008, p. 105). Inferring from their observations of 61 students participating in a university preparatory course, they suggest that it may be a promising strategy to unpack the conclusion by writing down a proof framework before getting started with the problem-centred part of a proof (Selden, J. & Selden, 1995). Other researchers emphasized the importance of analysing and using the structure of mathematical arguments in the context of learning proofs (Knipping, 2008; Pedemonte, 2007), as well.

Though all of these individual-mathematical activities and aspects may be useful for students to find a conjecture and to establish the validity of it, they do not necessarily lead to an interesting conjecture and an accepted proof. Until now, there is still limited knowledge about which activities actually predict the quality of the resulting proof.

7.3.2 Collaboration in mathematical conjecturing and proving

It is a common approach to treat proving as a cognitive activity employed by individuals with the aim to verify the correctness of a mathematical statement or to gain insight into why it is true is a common approach (e.g., Villiers, 1999). However, generating conjectures and developing proofs can also be considered from a social-discursive perspective. This is consistent with the view that conjectures are provided for the reflection of other mathematicians (and learners), sharing ideas and discussing arguments for or against them (Alibert & Thomas, 1991). What is accepted as a proof depends on the social context (Thurston, 1994) and on different criteria defined by the mathematical community (Stylianides, G. J., Sandefur, & Watson, 2016) and to be acquired in mathematical discourse. Within this perspective, proof might be seen as a “means of convincing oneself whilst trying to convince others” (Alibert & Thomas, 1991, p. 215). Participating in argumentative dialogues requires the ability to justify and explain a claim to peers on the one hand, as well as to interpret ideas of the speaker and to give feedback for correctness on the other hand. One challenge is to develop a “common ground” (Clark, H. H. & Schaefer, 1989). For instance, interacting with peers demands establishing a shared perception of what is recognized as a claim, an inferential rule or a given fact (Yackel & Cobb, 1996), building common frames of reference and resolving discrepancies in understanding (Barron, 2000). A very basic indication of successful collaboration is that students’ conversational turns built upon each other (e.g., Chi & Wylie, 2014; Roschelle & Teasley, 1995; Vogel et al., 2016). These so-called transactive (Teasley, 1997) or interactive (Chi, 2009) activities are attributed a high potential for fostering domain-general argumentation skills and deepening conceptual knowledge (Asterhan & Schwarz, 2009). According to Blanton and Stylianou (2014), interactive reasoning might also be seen as a discourse tool by which students can improve their proof understanding as well as their strategic knowledge in constructing proofs. Especially, criticizing or integrating the learning partner’s utterances are considered as interactive activities that are likely to trigger deep cognitive processes (Vogel et al., 2016). Therefore, the extent to which students monitor each other’s utterances, integrate divergent interpretations and, finally, make decisions together can be seen as an influential factor that may explain the variability of outcomes in collaborative math-problem-solving tasks (Barron, 2000; Clark, K., James, & Montelle, 2014).

Several studies have demonstrated that, on average, collaborative work may lead to better learning outcomes than individual engagement (cf. Barron, 2000; Cohen, 1994; Johnson, Johnson, & Smith, 2007), but this is not happening automatically. In unstructured learning environments (without any guidance), collaborators often tend to engage in low-level argumentation processes (e.g., Kollar et al., 2007), for instance, they rarely relate explicit evidence to their explanations (e.g., Sandoval, 2003). We conclude that students may benefit from collaboration and that learning occurs through interaction with peers when an atmosphere that enhances productive collaboration can be reached. Furthermore, we take the view that mathematical proving is at least partially a social activity and thus, are interested in understanding the relation between, and effects of, social-discursive and individual-mathematical characteristics of conjecturing and proving processes.

7.4 Process characteristics for collaborative conjecturing and proving

Research from different perspectives has put forward a number of theoretically plausible and partially empirically supported hypotheses about what constitutes mathematical conjecturing and proving processes that are successful in the sense that they lead to an interesting conjecture that goes well beyond the information available initially, and to a validly reasoned evaluation of this conjecture (e.g., Koedinger, 1998; Furinghetti & Morselli, 2009; Zazkis et al., 2015). However, a large proportion of these hypotheses has not been researched systematically so far, beyond mostly exploratory case studies. In the following, we will subsume them under the term *process characteristics of collaborative conjecturing and proving*. Our main assumption, based on the hypotheses put forward in the literature, is that collaborative conjecturing and proving processes that show an “ideal version” of one or more of these process characteristics will direct to interesting conjectures and valid proofs more frequently than those with a “low version” (ideal version: including all of the desirable aspects that ought to be present; low version: including none of these aspects). We differentiate between process characteristics that refer to primarily individual-mathematical aspects resp. social-discursive aspects (cf. Kollar et al., 2014).

7.4.1 Individual-mathematical process characteristics

In the following we present four individual-mathematical process characteristics referred from mathematics-educational research on conjecturing and proving. We agree with several researchers that the process of proving is complex, incorporating a wide range of activities such as identifying patterns, formulating conjectures, selecting given properties and structures, organizing logical arguments, and communicating them to a broader public – each of them is by no means trivial (e.g., Boero, 1999; Healy & Hoyles, 2000; Selden, A. & Selden, 2013a). The four individual-mathematical process characteristics describe the quality of how proving problems are explored, how precisely arguments are communicated, how different ideas are selected, combined and related to underlying mathematical concepts, and how claims are justified by data, warrants and modal qualifiers.

Cognitive complexity: Constructing mathematical proofs requires developing ideas that ideally show "why a particular claim is true" (Raman, 2003, p. 324) and that can be translated into arguments that demonstrate consistency to the reference theory. These "key ideas" combine two central aspects of conjecturing and proving, namely, (informal) conceptual understanding and a sense of formal rigour. For instance, using the symmetry property of even functions to argue that the derivative of an even function is odd, represents a key idea (Raman, 2003). Producing such kind of ideas may be done while generating examples or by finding suitable representations of definitions that provide access to the underlying logical structure and concepts (Alcock & Weber, 2010; Moore, 1994; Sandefur et al., 2013; Selden, J. et al., 2014). Yet, generating one's own examples and choosing appropriate definitions demand complex cognitive activities that go beyond simply replicating examples or definitions given by professors or textbooks (Moore, 1994). Forming analogies to related tasks, transferring concepts from one field to another or cross-referencing previous developed ideas seem to be cognitive activities that enable to gain new insights and to generate mathematical proofs successfully (Pease, A. & Martin, 2012). We conclude that developing different (key) ideas and bringing in new perspective in the proof construction process by searching for patterns, considering different cases, or changing flexible between formal and informal representation systems, as well as making explicit connection between ideas and mathematical concepts can be regarded as process characteristics that indicate the quality of conjecturing and proving processes.

Accuracy and precision of statements: Although mathematical ideas and proofs have to be evaluated with respect to a given context, they still must conform to the knowledge shared and accepted by the mathematical community. Furthermore, they need to be communicated in a "subject-specific, scientific language" (Engelbrecht, 2010). Accuracy and explicitness are considered as key aspects of successfully explaining mathematical ideas (Reinholz, 2016). The use of specific formal notations or symbols is certainly not necessary, but may facilitate the correct and precise communication of mathematical ideas and arguments. Wrong, inaccurate or implicit statements may lead the proof construction process on wrong paths and cause impasses and errors (Selden, J. et al., 2014). Savic (2015b) observed that mathematicians check, analyse and utilize incorrect proving steps more frequently than graduate students. Thus, capturing mathematical ideas and arguments in an accurate and precise way as well as identifying and rethinking incorrect assertions and impasses (Weber, 2009) form further promising process characteristics of conjecturing and proving processes.

Exploring and refining conjectures: The examination of an online discussion by Pease, A. and Martin (2012), where twenty-seven people from around the world solved a mathematical problem of International Mathematical Olympiad standard collaboratively, has shown that a large proportion of comments concerned conjectures. Participants proposed not only one, initial conjecture but also several sub-conjectures, explored their properties and limits, and expressed a level of confidence in them (Pease & Martin, 2012). Reformulating an initial conjecture and checking if the new reformulation is more transparent, or reflecting on how a counterexample would look like, seem to be promising strategies within conjecturing and proving (Weber, 2009). Especially when experienced mathematicians reach an impasse, they question the constraints of their conjectures and try to construct a counterexample (Savic, 2015b). Based on these observations, we assume that exploring and refining conjectures critically are an indication of deep cognitive processes and thus, represent activities that are essential for both generating conjectures and testing conjectures.

Argument structure: The importance of formulating structurally sound arguments has already been discussed in detail by several researchers (e.g., Toulmin, 1958; Krummheuer, 1995; Yackel, 2001) and constitutes a quality criterion for assessing argumentation in, for instance, online learning environments (e.g., Clark, D. B., Sampson, Weinberger, & Erkens, 2007). Formulating structurally sound arguments means that claims need to be based on some facts (called data) and that the legitimacy of the inferences connecting data and claim has to be explained by warrants and backings. Making implicit warrants explicit, as well as transforming data, claims and warrants of informal arguments into a more formal mathematical language contribute to the successful translation of informal arguments into verbal-symbolic proofs (Zazkis et al., 2016). Furthermore, Inglis, Mejia-Ramos, and Simpson (2007) emphasized that choosing adequate modal-qualifiers for each type of warrant play a crucial role within the proof construction process.

7.4.2 Social-discursive process characteristics

Analysing processes of collaborative argumentation is a central topic in educational psychology and in Learning Sciences research (especially, in CSCL research) (e.g., Asterhan & Schwarz, 2007, 2009; Kollar et al., 2014; Meier et al., 2007; Vogel et al., 2016; Weinberger & Fischer, 2006), as well. The motivation behind this research is to identify the challenges on the one hand and the aspects that are crucial for successful learning in collaborative environments on the other hand. Since engaging students in collaborative argumentative discourses is considered as an effective tool for enhancing students' understanding of challenging concepts (e.g., Clark, D. B. et al., 2007; Leitão, 2000), we also focus on social-discursive activities that refer to more general, dialogical argumentation skills. It is assumed that these skills are required for individuals to be able to participate successfully in mathematical debates. We define three social-discursive process characteristics that comprise the quality of how students contribute to a collaborative argumentative discourse by building upon and integrating each other's perspectives and ideas as well as by asking critical questions or proposing alternative strategies.

Critical questions: Clark, K. et al. (2014) conducted research on advanced calculus students working in small groups on problem solving tasks. The goal of their study was to examine the role of group interactions and to identify strategies that these students employ while problem solving. One of their results was that group members questioned each other for several reasons and during different phases, and that this was an important strategy to establish a collaborative atmosphere. Questions that prompt the learning partners to justify their approach as a specific type of transactive activities may lead to productive argumentation in collaborative settings (Teasley, 1997). Meier et al. (2007) also assumed that questions may contribute to a "smooth 'flow' of collaboration" (p. 82), and that they can be used for making sure to have the learning partner's attention.

Turn-taking sequences: The structure of turn-taking sequences is regarded as "as an indication of the degree to which students share common problem representations" (Roschelle & Teasley, 1995, p. 76). It is supposed that a shared conception of a given problem enables meaningful and profound conversations between peers. Interactive turn-taking sequences are characterized by turns that build upon each other, extend or challenge the ideas of others and take criticism and feedback of all learning partners into account (Chi, 2009; Roschelle & Teasley, 1995; Teasley, 1997). Establishing a shared conception of a subject matter includes integrating different perspectives, modifying initial arguments on the basis of the learning partner's contributions, and trying to make joint decisions (Weinberger & Fischer, 2006). We subsume all these activities that capture how students' turns reference to the contribution of others as a characteristic of productive collaboration.

Reacting to the learning partner's errors: In collaborative settings, learners often tend to accept the ideas of their learning partners without being critical about what they have said ("quick consensus building"; Weinberger & Fischer, 2006, p. 84). Yet, scrutinizing the information provided by others is crucial for being able to participate successfully in collaborative argumentations. Identifying invalid arguments that contain fallacies or errors in reasoning constitutes an important characteristic of high-quality argumentation (Mayweg-Paus, Thiebach, & Jucks, 2016). To address the mistakes of the learning partner's arguments means expressing criticism. Since conflict-oriented consensus building has been put forward as an important aspect in collaborative settings, learners need to be explicit about the errors they have identified, modify them or present alternative approaches (Weinberger & Fischer, 2006). This characteristic combines individual-cognitive and social-discursive features of argumentation, in the way that an adequate reaction to the learning partner's errors requires domain-specific knowledge, accuracy and precision (e.g. Reinholz, 2016), as well as the social act of monitoring and evaluating other people's verbal and written utterances (Pease & Martin, 2012; Vogel et al., 2016).

Taken together, we suppose that important aspects of conjecturing and proving in collaborative situations can be measured along these seven process characteristics. Based on the literature, we further hypothesize that processes that present almost the "ideal version" of each of these seven properties are likely to lead to an interesting conjecture resp. an acceptable proof during the collaboration, and thus reflect the quality of collaborative conjecturing and proving processes to a large extent. This set of process characteristics is certainly not exhaustive. For example, meta-cognitive activities are currently not considered explicitly. However, the preliminary collection provides a first attempt to structure a subset of process characteristics of collaborative conjecturing and proving, and to investigate them empirically.

Since symbol schemes only code whether or not a certain mode of behaviour occurs and category schemes only assign the observed characteristic to a specific category disregarding its quality (Seidel, Prenzel, & Kobarg, 2005), rating scales that allow to capture the quality of a characteristic seem to be an appropriate method for assessing conjecturing and proving processes along the seven process characteristics we have defined. In terms of the degree of inference required for the coding, Seidel et al. (2005) classified the directly observable symbol scheme as low inference, the category scheme as middle inference, and the rating scales, for which a certain amount of interpretation is necessary, as high inference coding schemes. Low inference coding schemes (such as checklists) are often criticized for trivialization. Usually, they are not able to reflect one's behaviour and performance on a task and thus, the underlying construct (Newble, 2004). In the educational domain, it has been shown that complex constructs and structures can be measured more validly with high inference rating schemes (e.g., Seidel et al., 2005; Gartmeier et al., 2015). Consequently, we follow the approach of creating a high inference rating scheme consisting of seven rating scales presenting the identified process characteristics of conjecturing and proving.

7.5 The current study

Even though students' proving difficulties have been frequently discussed (e.g., Selden, A. & Selden, 2008; Selden, A. & Selden, 2011; Ufer et al., 2008; Weber, 2001), there is still little knowledge about how good collaborative conjecturing and proving processes could be described and measured. The current study presents a high inference rating scheme that facilitates a systematic analysis of students' collaborative conjecturing and proving processes in an open-ended conjecturing task from an individual-mathematical and social-discursive perspective. Based on our systematization of process characteristics from the literature, we were interested in (1) if and how these process characteristics of collaborative conjecturing and proving can be measured reliably. This question addresses the fact that the assessment of students' collaborative conjecturing and proving processes along different characteristics should be consistent between different observers. To approach this question, independent coders rated a set of video-taped collaborative conjecturing and proving processes for an open conjecturing and proving task independently. Inter-rater correlations as values for the inter-rater agreement were used to determine the extent to which the two raters came to the same classification of quality levels for each process characteristic (Seidel et al., 2005). Intraclass-correlations that estimate the inter-rater agreement by comparing the variability of different ratings of the same observer to the total variation across all ratings and all observers were calculated, as well (Uebersax, 2010). We expected (1.1) a satisfactory inter-rater agreement (ICC $\geq$ 0.6). This cut-off point was chosen in accordance with existing guidelines (Cicchetti, 1994; LeBreton & Senter, 2008). Moreover, and more exploratory, we aimed (1.2) to study whether the members of a dyad would contribute equally to the quality of collaborative conjecturing and proving processes, or if the scores for two members of a dyad would be rather independent from each other. From a methodological perspective, these data can inform future studies about the adequate level of analysis (individual vs. group) when considering similar collaboration processes.

Furthermore, we were interested in (2) the empirical structure of the observed process characteristics. In particular, we assumed (2.1) that individual-mathematical and social-discursive process characteristic of collaborative conjecturing and proving would each form a consistent dimension. Prior studies have shown low correlations between individual-mathematical and social-discursive aspects of mathematical proof skills (Kollar et al., 2014). Thus, and based on how we conceptualized individual-mathematical and social-discursive process characteristics, we expected (2.2) low correlations between the two dimensions.

7.6 Method

7.6.1 Setting and sample

The current study was embedded in a voluntary two-week preparatory course for prospective mathematics university students. The course contained lectures and seminars about elementary number theory and other basic mathematical topics (e.g. proof techniques, propositional and predicate logic, elementary set theory). $N = 98$ undergraduate mathematics students (M_{age} = 19.83, SD_{age} = 3.56; 44 female, 3 without specification) participated in this study on the sixth day of the course. Students worked in dyads in a computer-supported collaborative learning environment on the following open-ended conjecturing and proving task: *"Take four consecutive numbers, multiply them, and add one. Repeat this and try to find similarities. Formulate a conjecture and prove it!"*. We tried to keep the content knowledge required to solve the problem at a relatively low level, so that the participants' reasoning would not be restrained by their lack of content knowledge to a large extent

Each student was provided with one laptop and one graphics tablet that allowed them to exchange written ideas, visualisations or arguments within their dyad using a graphical chat. The students were also allowed to communicate to each other verbally face-to-face. All screen and audio activities were recorded by the laptops and transformed into a video file. More details about the design of the computer-supported learning environment can be found in Kollar et al. (2014) and Vogel et al. (2016).

Before starting collaboration, the learners received a short introduction (around 7 min) about how to operate with the technology and about how to work collaboratively on this type of task. To avoid specific effects of grouping peers with substantially different prior achievement (Webb et al., 2002) on students' conjecturing and proving processes, the students were assigned to homogenous dyads with respect to prior school achievement. The collaboration session took between 45 and 60 minutes.

7.6.2 Coding manual and rating procedure

To quantify the quality of students' conjecturing and proving processes, we developed a theory-based rating scheme that incorporated high inference ratings for the individual-mathematical and social-discursive process characteristics introduced above. Each rating scale comprised five quality levels. For each of seven process characteristics the rating scheme provided a detailed definition, a description of the "ideal version" including desirable aspects that ought to be present as well as undesirable aspects that ought to be absent, illustrative examples, and coding rules. We labelled and operationalized the process characteristics as:

(1) *Cognitive complexity* focuses on how key ideas are developed and combined at different phases of the proving process. (2) *Accuracy and precision* is the extent to which mathematical arguments and ideas are captured correctly and precisely. (3) *Critical exploration* refers to the extensive investigation of conjectures by generating examples and counter-examples, exploring the constraints of a conjecture, and by formulating more than one initial conjecture. (4) *Argument structure* measures whether mathematical claims are based on data and whether appropriate warrants are provided that explain the legitimacy of the data. (5) *Critical questions* refers to utterances that critically questions the ideas of the learning partner or one's own

solution steps and demand an explicit answer of the learning partner. (6) *Turn-taking-sequence* includes to what extent the learner takes the partner's contribution into account either by extending the learning partner's ideas or by using the same approach. (7) *Reaction to the learning partner's errors* scores to what extent the learner identifies errors and impasses produced by the learning partner and whether the learner explains these errors or proposes alternative solution steps.

All process characteristics were judged on a 5-point Likert scale with lower values indicating better performance and high values indicating lower quality (comparable with the German school grade system) (Gartmeier et al., 2015). To minimize the complexity of these global judgments, we divided the whole working process of each dyad into two parts of equal length. The first half of the video was coded along the seven process characteristics, separately for each learner. Afterwards, the second half was rated in the same way. Thus, the conjecturing and proving processes of each learner were assessed twice along each of the process characteristics. An excerpt from the coding scheme to illustrate the operationalization of the process characteristic *cognitive complexity* can be found in Figure 9. When most of time was spent with off-task talk and hence, the coding of a specific process characteristic was not possible, we coded it as missing value.

7.6.3 Rater Training

It is well known that the inter-rater reliability of high inference ratings is often lower than in more standardized assessment methods such as low inference coding systems (Seidel, 2005). Since applying high inference ratings challenges the observers to interpret students' activities and to make decisions that go beyond the directly observable behaviour (Herweg, Seidel, & Dalehefte, 2005), an intensive rater-training was needed. The two raters participated in a three-day training to get an introduction to the theoretical concepts behind each process characteristic, to obtain a sound understanding of the coding dimensions and levels of the rating scales, and to convey the expected range of performance and the corresponding scores (cf. Langer & Schulz von Thun, 2007). The training consisted of six sessions (spread over three days), which included activities such as sorting videos according to the quality of students' conjecturing and proving processes. Several proof attempts were evaluated to develop a shared understanding about which arguments have to be supported by warrants and backing, and which key ideas have to be formulated and combined to construct an acceptable proof. Other sessions included discussions about common rater-errors and strategies for avoiding them. Three more training phases followed in which videos were first coded by each rater individually and then discussed together in the group. The training videos came from a previous study and were excluded from further analyses.

Question: To what extent does the learner generate new ideas, establish logical relationships between different ideas and concepts, and thus, introduce new aspects at all relevant phases?

1	The learner generates new ideas at all different phases of the conjecturing and proving process and relates most of them to each other. Given elements and previously constructed concepts and theorems are combined in new ways. In all appropriate situations, the learner brings in new perspectives on the conjecturing and proving process by considering different cases, identifying patterns and logical relationships between them, or by changing the representation system. A global view on the conjecturing and proving process is consistently observable.
2	The learner generates new ideas at many different phases of the conjecturing and proving process and relates some of them to each other. Given elements and previously constructed concepts and theorems are sometimes combined in new ways. In many appropriate situations, the learner brings in new perspectives on the conjecturing and proving process by considering different cases, identifying patterns and logical relationships between them, or by changing the representation system. A global view on the conjecturing and proving process is largely observable.
3	The learner generates new ideas at some different phases of the conjecturing and proving process and relates only a few of them to each other. Given elements and previously constructed concepts and theorems are sometimes combined. In some appropriate situations, the learner brings in new perspectives on the conjecturing and proving process by considering different cases, identifying patterns and logical relationships between them, or by changing the representation system. A global view on the conjecturing and proving process is only partly observable.
4	The learner generates only a few new ideas and doesn't relate them to each other. Given elements and previously constructed concepts and theorems are only seldom combined. The learner mostly repeats or paraphrases existing ideas or already known concepts/ approaches. A global view on the conjecturing and proving process is rarely observable.
5	The learner generates only a few ideas and doesn't look beyond single aspects. A global view on the conjecturing and proving process is never observable.

Figure 9: Quality levels of the first process characteristic "cognitive complexity"

7.7 Results

7.7.1 Descriptive results and reliability of the coding scheme

First, we investigated (1.1) whether the two trained observers achieved a satisfactory inter-rater agreement. After the two raters had analysed 20% of the sample with the rating scheme independently from each other, intra-class correlations and inter-rater correlations were calculated for all ratings (see Table 7). The inter-rater correlations (M_r = .87, SD_r = .11) as well as the intra-class correlations (M_{ICC} = .90, SD_{ICC} = .11) with values higher than the set-cut off point (ICC $\geq$ .60) showed that, for a high inference rating scheme, the observed process characteristics led to highly reliable outcomes (Cicchetti 1994; LeBreton & Senter, 2008). Table 8 provides the mean values and standard deviations for each process characteristic, separately for the first and the second half of the collaborative working process. Across all process characteristics, the incoming university mathematics students achieved a mean value of 2.93 (*SD* = 0.52). The mean values of the process characteristics ranged from 2.13 to 3.67, and a slight decrease in the mean values from the first to the second half of the collaboration working process was observable for most of the process characteristics. The mean values of the process characteristics *critical exploration* and *reaction to the learning partner's errors* tended to be lower compared to the mean values of the other process characteristics. *Accuracy and precision* and *turn-taking-sequence* were the process characteristics with the highest mean values at both measurement intervals. Ceiling or floor effects could not be observed for any of the process characteristics. The standard deviations indicated that there was a substantial dispersion between the individuals.

To analyse (1.2) whether the scores for two members of a dyad were more similar to each other than to the scores for participants of other dyads, we compared the variance within dyads with the variance between dyads (ρ), in the ratings of each process characteristic using intraclass-correlations (see Table 9). Results showed small values for ρ for the process characteristics *cognitive complexity*, *critical questions* and *turn-taking-sequence* for the first half of the collaborative working process, implying considerable differences between students within a dyad across these three characteristics. The values for ρ for the process characteristics *accuracy and precision*, *critical exploration*, *argument structure* and *reaction to the learning partner's errors* were already substantially within the first half. For the second half of the collaborative working process, the variance within the dyads was much smaller than the variance between the dyads for all process characteristics, indicating a more similar contribution of the students in a dyad to the quality of the collaborative conjecturing and proving process later in the working process. Note that the values for ρ, describing the similarity of students' contributions to the process quality, did not increase for the process characteristic *reaction to the learning partner's errors* (see Table 9). The willingness of the students to critically examine the arguments of their counterparts remains constantly low.

Table 7: *Inter-rater correlations and intra-class correlations for all process characteristics.*

	First half of the collaboration		Second half of the collaboration	
	Inter-rater correlations (r_s)	Intra-class correlations (ICC)	Inter-rater correlations (r_s)	Intra-class correlations (ICC)
Cognitive complexity	.960**	.966	.923**	.901
Content accuracy	.977**	.969	.967**	.995
Critical exploration	.890**	.968	.985**	.986
The soundness of arguments	.675**	.922	.902**	.903
Critical questions	.836**	.826	.813**	.942
Turn-taking-sequence	.766**	.749	.914**	.961
Reaction to the learning partner	.958**	.968	.681**	.618

Table 8: *Mean values and standard derivations for all process characteristics.*

	First half of the collaboration M (SD)	Second half of the collaboration M (SD)
Cognitive complexity	2.612 (.970)	3.109 (1.262)
Content accuracy	2.163 (.938)	2.341 (1.147)
Critical exploration	3.367 (1.255)	3.637 (1.269)
The soundness of arguments	3.052 (.863)	3.000 (1.206)
Critical questions	2.663 (.907)	3.121 (1.172)
Turn-taking-sequence	2.133 (.857)	2.596 (1.276)
Reaction to the learning partner	3.540 (1.305)	3.672 (1.300)

Table 9: *Intra-class correlations as values for the between-cluster variance.*

	First half of the collaboration ICC	Second half of the collaboration ICC
Cognitive complexity	.000	.382
Content accuracy	.425	.620
Critical exploration	.347	.749
The soundness of arguments	.176	.622
Critical questions	.000	.216
Turn-taking-sequence	.011	.711
Reaction to the learning partner	.384	.276

7.7.2 Structure of individual-mathematical and social-discursive process characteristics

In the following analyses, we only included the data from the first half of the collaborative working process due to a larger amount of missing values in the second half. These missing values in the second half of the collaboration processes are attributable to the fact that some dyads stopped to work on the proving task earlier and started to have private conversations.

To investigate the empirical structure of the process characteristics, we performed confirmatory factor analyses for categorical data with M*plus* 7.4 (Muthén & Muthén, 2015). We took the hierarchical structure of the data (individual students nested within dyads) into account by using the “TYPE=COMPLEX” option of the “ANALYSIS command” (Muthén & Muthén, 2015). Results indicated that the two-dimensional model with one individual-mathematical and one social-discursive component showed good fit indices ($\chi^2(13) = 20.61$, $p = .08$, RMSEA = .07, CFI = .95, WRMR = .63). The factor loadings were significantly different from zero for all process characteristics, except for the process characteristic *reaction to the learning partner’s errors* (see Table 10).

In order to improve the two-dimensional model fit, we reconsidered this process characteristic and decided to treat it as an individual-mathematical one (a theoretical justification of this decision is provided in the discussion section that follows) and conducted a further factor-analysis with four individual-mathematical and two social-discursive process characteristics. This model showed a very good fit ($\chi^2(13) = 13.14$, $p = .44$, RMSEA = .01, CFI = .99, WRMR = .47) and substantial factor loadings for all process characteristics (see Table 4). The individual-mathematical and social-discursive factors were not significantly correlated, and consequently, the one-dimensional model did not fit the data adequately well ($\chi^2(14) = 63.72$, $p<.001$, RMSEA = .19, CFI = .639, WRMR = 1.21). These findings as well as the theoretical foundation support our assumptions (2.1) that individual-mathematical and social-discursive process characteristics of collaborative conjecturing and proving can be treated as coherent constructs in themselves, and (2.2) that these two dimensions should be distinguished when analysing collaborative conjecturing and proving processes with our instrument.

Table 10: *Factor loadings for both 2-dim model.*

	First model		Second model	
	First factor (individual-mathematical)	Second factor (social-discursive)	First factor (individual-mathematical)	Second factor (social-discursive)
Cognitive complexity	.770**		.758**	
Content accuracy	.321**		.338**	
Critical exploration	.568**		.576**	
The soundness of arguments	.684**		.686**	
Critical questions		-.873**		.768**
Turn-taking-sequence		-.706**		.801**
Reaction to the learning partner		.048	.303**	

7.8 Discussion

Conjecturing and proving are highly valued mathematical practices to demonstrate students how knowledge is dynamically evolving within the discipline of mathematics (Komatsu, 2016). In the past, conjecturing and proving were investigated either from an individual-mathematical or social-discursive perspective. Mathematics educational researchers have mainly focussed on mathematics-related conjecture and proof aspects. For instance, Yang (2012) evaluated the quality of how students formulated conjectures, and found that successful students guided their thinking by the use of examples and by recognizing the logical structure of an argument. Other researchers investigated the transition from informal arguments to formal proofs (Gibson, 1998; Zazkis, et al., 2016), or categorized the structure of students' argumentations and proofs (e.g., Inglis, et al., 2007). We summarized those aspects under the term individual-mathematical process characteristics. Research in educational psychology and in the Learning Sciences proposed different strands of research regarding the social aspects of argumentation. Studies that took those aspects of argumentation, that we call social-discursive process characteristics, into account investigated how students can benefit from the interaction with peers in collaborative learning situations, and how they can develop a common understanding of a problem (e.g., Asterhan & Schwarz, 2007; Meier et al., 2007; Roschelle & Teasley, 1995; Vogel et al., 2016). In this study, we attempted to include both perspectives, we systematized process characteristics of collaborative conjecturing and proving that have been extensively discussed in the literature, and we developed a rating scheme that incorporates these theory-based process characteristics. We evaluated the new rating scheme by analysing the collaborative conjecturing and proving processes of 49 dyads and investigated the empirical structure of the observed process characteristics.

7.8.1 Summary of the results

Regarding the reliability of measuring collaborative conjecturing and proving processes, our results showed that the inter-rater agreement across all high inference ratings was very good. Achieving a satisfactory consistency between different raters is challenging with high inference rating scales in general (Newble, 2004; Seidel, 2005). Making qualitative judgments that include interpretations that go beyond directly observable behaviour often leads to a reduced reliability of the ratings (Seidel, 2003). As in prior studies, the intensive rater training proved crucial to obtain reliable data. From a theoretical point of view, the advantages for the development of high inference ratings are given by their validity (e.g., Newble, 2004). Until now, high inference rating schemes have been used more often in learning climate research (Clausen, 2002) and hence, in the area of educational science (e.g., Seidel et al., 2005) than in mathematics educational research. However, the new analytic rating scheme allowed us to capture seven theory-based process characteristics of conjecturing and proving (four deduced from mathematics educational research and three from research in educational psychology and the Learning Sciences) reliably.

In a next step, we analysed how similarly the learners of one dyad contributed to the quality of collaborative conjecturing and proving processes. Our results indicate that the within-dyad similarities seem to be more pronounced for some process characteristics (*accuracy and precision*, *critical exploration, argument structure* and *reaction to the learning partner's errors*) than for others (*cognitive complexity, critical questions* and *turn-taking-sequence*). This points out that some aspects seem to align during the collaborative process, while others do not. One reason for a missing alignment might be students' different (learning) prerequisites (e.g., complementary knowledge that they cannot integrate) or different expectations about collaboration (cf. Meier et al., 2007). Moreover, the relatedness of students working within one dyad increased from the first to the second half of the collaborative working process for most of the process characteristics. This is in line with the findings of other researchers that students in dyads seem to imitate each other increasingly over time and thus, develop similar behavioural patterns (Anjewierden, Gijlers, Kolloffel, Saab, & Hoog, 2011). Furthermore, we observed a slight decrease in the quality of how individual-mathematical and social-discursive activities are employed during the working process from the first to the second part of collaboration. One explanation for this might be that students lost motivation due to the complexity of the task, and thus were less engaged in producing new ideas or validating a completed argument after some time. Furthermore, they may lack confidence in their own abilities to develop something new or something better than before (Selden, A. & Selden, 2013a).

Our assumption about the consistency of the individual-mathematical respectively social-discursive component of collaborative conjecturing and proving processes could be confirmed (with one exception, see below), as well as that both dimensions can be empirically distinguished from one another. The individual-mathematical component takes into account the quality of exploring and formulating conjectures (e.g., Koedinger, 1998; Pease, A. & Martin, 2012) as well as the quality of generating accurate (e.g., Reinholz, 2016) and structurally sound arguments (e.g., Toulmin, 1958) that are related to the underlying concepts. This component mainly refers to domain-specific knowledge that is necessary, for instance, to develop a mental representation of the problem-situation and to identify crucial properties that offer an access to deduce further inferences. Moreover, the process-characteristics that are related to this component may differ across domains. For example, transferring the aspects for generating structurally sound arguments to other disciplines appears problematic as, in mathematics, the only acceptable evidence is a chain of deductive arguments, while other sciences may allow empirical-inductive methods as well. The social-discursive component describes more domain-general argumentation skills such as questioning (e.g., Mayweg-Paus et al., 2016) or refining the learning partner's contributions (Teasley, 1997). These process-characteristics may occur in a similar manner across different domains.

In comparison to the initial model with four individual-mathematical and three social-discursive process characteristics, an adapted model that categorized the characteristic *reaction to the learning partner's errors* as an individual-mathematical one fitted the data better. This might be mainly due to the fact that the identification of errors and impasses is strongly related to domain-specific knowledge of mathematics and additionally, constitutes the prerequisite for all the further activities that make up this process characteristic such as explaining the error to the learning partner.

Regarding the empirical structure of the observed process characteristics, our results are consistent with findings from prior studies that the individual-mathematical and social-discursive components of mathematical argumentation are widely unrelated to each other. The measurement of the students' underlying individual-mathematical and social-discursive argumentation skills, using written test items, revealed a low correlation between the two facets

(Kollar et al., 2014). With respect to fostering mathematical argumentation skills, the meta-analysis by Vogel, Wecker, Kollar, and Fischer (2017) indicates that supporting social-discursive aspects of mathematical argumentation skills is not necessarily effective, unless it is accompanied by domain-specific support addressing individual-mathematical skills or process characteristics. We conclude, both facets of mathematical argumentation processes are important, even though they showed no substantial correlation.

In sum, this study provides first evidence for the feasibility of measuring theory-based process characteristics of collaborative conjecturing and proving: The good interrater-agreement as well as the internal structure of our data, which largely fits our theoretical assumptions, indicate that the quality of students' collaborative conjecturing and proving processes can be categorized and evaluated by applying our rating scheme.

7.8.2 Limitations of the study

Despite these promising results, some limitations of the study need to be acknowledged. First of all, participating in the preparatory course in which the study was embedded was voluntary. Self-selection effects may have led to a selective sample, e.g. with high interest in mathematics. Furthermore, our sample consisted only of prospective mathematics university students before the beginning of their first semester. Due to the selectivity of the sample, our results have to be generalized carefully. It might be that graduate students or expert mathematicians focus more on the arguments and conceptual issues proposed by their counterpart (e.g. explicitly express their disagreement or formulate counter-arguments) and thus, their social-discursive activities can be more closely related to their individual cognitive thinking (Leitão, 2000). Moreover, we applied our analysis scheme for students' collaborative conjecturing and proving processes to one task in this study. Task-specific factors that may influence the results cannot be eliminated, and are an important field for further research. In the future, proof processes for several tasks might be compared to determine the internal structure of the process characteristics of collaborative conjecturing and proving (cf. Blömeke et al., 2015). Finally, the small number of process characteristics especially for the social-discursive component is a further limitation of the present study. The social-discursive component may comprise a broader range of process characteristics, such as the co- and socially shared regulation activities (Järvelä & Hadwin, 2013). The individual-mathematical set of process characteristics may be expanded as well by adding additional characteristics such as the quality of the use of formal-symbolic notations (Ottinger, Kollar, & Ufer, 2016). To adapt our coding scheme to other contexts, the specific rating rules and scale's anchors will have to be modified to the features of the collaborative conjecturing and proving situations one intends to evaluate. Since this study was a first attempt to measure collaborative conjecturing and proving processes from an individual-mathematical and social-discursive perspective, the set of process characteristics might not be exhaustive. Yet, it merely aimed at demonstrating how different process characteristics of collaborative conjecturing and proving might be operationalized and observed in students' behaviour. Finally, it still remains an open question whether these process characteristics that are assumed to be crucial for the success indeed predict the quality of the resulting outcome.

7.8.3 Conclusion and directions for future research

This study extends prior research in several ways: (1) It systematizes several individual-mathematical and social-discursive process characteristics from previous studies that have been reported in combination with successful conjecturing and proving, or argumentation outcomes. (2) It provides a rating scheme for assessing students' collaborative conjecturing and proving processes. High inference rating strategies that worked effectively in the medical or educational domain (Gartmeier et al., 2015; Seidel, 2005) were successfully adapted and transferred. (3) Substantial standard deviations, the high inter-rater agreement as well as the consistency within the two dimensions showed that our approach is feasible in principle: The developed rating-scheme comprises seven process characteristics of collaborative conjecturing and proving that could be reliably measured and analysed with regard to the within-dyad-similarities and the internal structure of the observed characteristics. Furthermore, our rating scheme allows for a direct assessment of collaborative conjecturing and proving processes, since it does not require the transcription of students' dialogues. It is time efficient and, from an educational perspective, it may be used for instructional purposes. Tutors may use this rating scheme, if they are taught how to apply it to identify where support is needed and which aspects they have to encourage most. As the rating scheme is largely content-neutral, it may be adapted to a variety of tasks. (4) Regarding the within-cluster variance we observed relatively high values for some process characteristics pointing out that it is necessary to consider each individual's contributions separately, not only at the dyad level, when evaluating collaborative conjecturing and proving processes. (5) On the other hand, for most of the process characteristics, especially at the second half of the collaborative working process, the within-cluster variance was quite low. Analyses that do not take this clustering into account may result in underestimation of standard errors and overestimation of statistical significance (cf. Lee, V. E., 2000). (6) Finally, this study investigated the empirical structure of individual-mathematical and social-discursive process characteristics of collaborative conjecturing and proving. Results indicated that collaborative conjecturing and proving processes can be conceptualized as a two-dimensional construct. As a practical consequence, in order to encourage students' collaborative conjecturing and proving skills, it would be useful to design learning environments that provide support for both components (cf. Vogel, et al., 2017).

The findings of this study delivered empirical evidence that our rating-scheme constitutes a valuable instrument for analysing students' collaborative conjecturing and proving processes from an individual-mathematical and social-discursive perspective. It has potential for further usage in research and teaching. Since the validity criterion plays an important role in the development of instruments (e.g., Blömeke, et al., 2015), future studies should investigate which of the process characteristics we have extracted from the literature actually predict the quality of the resulting product of collaborative conjecturing and proving processes. Moreover, the rating scheme may be applied to detect effects of interventions or scaffolds that systematically foster one of the two quality facets. Making it useable for practitioners could also be a next step, as the rating scheme may support instructors and tutors while monitoring and supporting students' proof processes. It may be used to help tutors or lecturers learn to notice and interpret important characteristics of students' collaborative conjecturing and proving processes and thus, to enhance their "professional vision" for these processes (Goodwin, 1994; van Es & Sherin, 2002).

8 Study III

Generating structurally sound and accurate arguments: Key characteristics of successful collaborative conjecturing and proving processes

8.1 Abstract

There seems to be general consensus on the assumption that specific process characteristics of conjecturing and proving are crucial for their success, even though they may not emerge directly in the final product (the formulated conjecture and the constructed proof). Based on the literature, we have selected four individual-mathematical and three social-discursive process characteristics of collaborative conjecturing and proving that are considered essential for the successful production of conjectures respectively proofs. We empirically investigated to which extent these characteristics predicted the quality of the final conjecture and proof. Furthermore, we were interested in studying their relations to students' prior knowledge on proof. Therefore, we examined the interaction of *N=98* prospective mathematics university students working collaboratively on a conjecturing and proving task. Results indicate that generating structurally sound and accurate arguments during the collaborative discourse are key characteristics of successful conjecturing and proving processes. Furthermore, this study shows that individual-mathematical process characteristics mediate the relation between students' prior knowledge on proof and the quality of their resulting conjectures and proofs. We present more detailed analyses of the process characteristics and their effects on the specific quality aspects of the final product and discuss implications for teaching and research.

8.2 Introduction

Inquiring mathematical conjectures, finding supporting arguments, and discussing them with peers are the daily work of mathematicians and hence considered as a substantial objective of mathematics education (e.g., Mariotti, 2006; Stylianides, A. J., Bieda, & Morselli, 2016). However, developing these complex skills is challenging for most students (Heinze et al., 2005; e.g., Heinze & Reiss, 2004; Moore, 1994; Selden, A. & Selden, 2008). In other disciplines such as politics, psychology, or philosophy, students also have to be able to generate evidence-based arguments.

This might be one reason why there is a widespread trend towards describing and analysing students' behaviour when solving argumentation tasks. Researchers from several disciplines focus their attention on investigating what the crucial aspects of scientific reasoning and argumentation are (Fischer et al., 2014). Learning more about this may help to discover the causes for students' main difficulties and to design adequate scaffolds. Of course, it is possible to come up with characteristics of "good" argumentation processes by theoretical analyses or by observing students' argumentative activities. However, the success of a good conjecturing and proving process is determined by its outcome, the final conjecture respectively proof.

In this contribution, we are interested in the relations between characteristics of the processes, the final conjecture and proof, and the students' prior knowledge on proof (representing their individual (learning-) prerequisites). Along with other researchers, we conceptualized conjecturing and proving processes as specific types of argumentative activities (e.g., Selden, A. & Selden, 2013b) and differentiate between an individual-mathematical and social-discursive component of argumentation (Kollar et al., 2014). We reviewed literature from mathematics education and educational psychology as well as the Learning Sciences to find potentially predictive process characteristics of collaborative conjecturing and proving.

There are several motivations behind our approach to investigate the relations between the processes, the final product, and students' prerequisites. First, analysing the relations between process characteristics and the resulting product may allow to uncover key characteristics for success. Secondly, it might help to detect specific challenges of students' and thereby to find out where support is required. Based on this knowledge, new scaffolds may be designed. If the relation is due to general (learning-) prerequisites, developing scaffolds with the focus on the processes may become less important. Furthermore, we expect to receive information that may be used for diagnostic purposes to identify dyads that will run into trouble. One may also be interested in studying the relations between process characteristics and prerequisites in order to be able to diagnose more general traits such as prior knowledge. However, the main objective of this study was to investigate the effects of individual-mathematical and social-discursive process characteristics on the quality of the resulting conjecture respectively proof by controlling for prior knowledge on proof.

The first part of this paper describes different process characteristics of collaborative conjecturing and proving and hence provides an overview about what counts as good collaborative conjecturing and proving processes from a theoretical point of view. In the second part, we will present the results of an empirical analysis capturing the relations between different individual-mathematical and social-discursive process characteristics, the quality of the resulting conjecture and proof (as the outcome of students' collaborative conjecturing and proving processes), and their prior knowledge on proof.

8.3 Research on conjecturing, proving and argumentation

In the past, the concepts of argumentation and proof have received serious attention by numerous researchers in mathematics education (Stylianides, G. J. et al., 2017; e.g., Stylianou, Blanton, & Knuth, 2010). While there was some debate on the relationship between argumentation and proof (e.g., Balacheff, 1999; Garuti, Boero, Lemut, & Mariotti, 1996;

Krummheuer, 1995), there is no disagreement that the consideration of both concepts is important to include all processes that are crucial in the context of poof (Mariotti, 2006). Since different definitions of the terms *argumentation* and *proof* are pursued by different research traditions (Balacheff, 1999; Stylianides, A. J. et al., 2016), we delineate our perspectives on conjecturing, proving, and argumentation in the following section. Furthermore, we theorize what set of skills are needed to solve conjecturing and proving tasks.

8.3.1 The relation between conjecturing, proving and argumentation skills

Argumentation is often described as an activity that mathematicians engage in when constructing proofs or conjectures, involving a variety of processes such as working with examples or using rhetorical tools to convince others that a statement is true. However, not every mathematical argument constitutes a proof (Selden & Selden, 2013), and not every argumentation task requires the generation or refinement of a conjecture. Consequently, we conceptualize proving and conjecturing as specific types of argumentation. Conjecturing comprises the activity of formulating a conjecture “according to some given information which could include either ill-defined or well-defined problems” (Lin, F. L. et al., 2012, p. 309). The formulation of a conjecture often results from exploratory activities, such as generating examples, testing a variety of cases, and attempting to identify patterns between them (e.g., Ellis et al., 2017; Koedinger, 1998). By proving, we are referring to the activity that a student or mathematician engages in to produce a connected sequence of logical inferences (i.e., proof), which satisfies the sociomathematical norms (Yackel & Cobb, 1996) and values so that the sequence of inferences will be acceptable to his/ her peers (Dawkins & Weber, 2017). Furthermore, we build upon the hypothesis that there exist some kinds of continuity, conceptualized as *cognitive unity*, between conjecturing and proving processes (cf. Garuti et al., 1996). The construction of a logical chain of inferences may be more ‘accessible’ to students when they get involved with the exploratory activities related to the formulation of a conjecture (Pedemonte, 2008).

Mathematicians have to negotiate with their colleagues about which conjectures are interesting for further explorations, which argumentations are accepted as proofs, and which pieces of their arguments have to be specified explicitly (e.g., Inglis & Aberdein, 2014; Manin, 1977). These definitions of conjecturing and proving merge individual and social point of views. The individual-mathematical perspective encompasses the ability to formulate conjectures, to select theorems that are related to the given problem, to generate appropriate and valid arguments, and to combine them to a deductive proof or refutation (e.g., Boero, 1999; Koedinger, 1998). To acquire this complex ability, students need a deep conceptual understanding in the area of the proof problem, knowledge about the nature of mathematical arguments (e.g., Ufer et al., 2008), strategic knowledge (Sommerhoff, 2017; Weber, 2001) as well as problem solving skills (e.g., Schoenfeld, 1992). The ability to participate in argumentative discussions, to debate about the acceptability of arguments, to justify one's own proving steps, to take the arguments of others into account, and to synthesize different contributions constitute the more domain-general, social-discursive dimension of mathematical argumentation (e.g., Kollar et al., 2014). Generally, it is assumed that social-discursive argumentation skills play a crucial role in collaborative knowledge construction and problem-solving activities (e.g., Kollar et al., 2007). In this project, mathematical argumentation is understood as a complex cognitive skill that includes both dimensions, one genuine to mathematics and the other related to the more domain-general, social-discursive component of argumentation.

8.3.2 Mathematical argumentation as a complex cognitive skill

Current discussion in research emphasizes the importance of conceptualizing skills and pursues the question of how skills become visible and accessible (e.g., Koeppen et al., 2008). Inspired by the framework proposed by Blömeke et al. (2015), we consider argumentation skills as a latent construct that becomes visible in the processes and the performance of an individual in a proof-related situation. From this perspective, conjecturing and proving processes as well as the performance can serve as indicators for an individual's mathematical argumentation skills. The performance can be observed and evaluated in the light of pre-defined criteria that are conform to the sociomathematical norms (cf. Yackl & Cobb, 1996). Furthermore, is assumed that the quality of these processes and the performance is affected by the individual's cognitive (learning-) prerequisites, such as prior knowledge on proof. Conjecturing and proving processes can be viewed as connecting elements that lead from the cognitive prerequisites on the one hand to the situation-specific performance on the other hand. To what extent conjecturing and proving processes mediate between one's cognitive prerequisites and performance has to be investigated empirically.

Examples of different situations requiring argumentation skills are constructing, reading, or presenting conjectures and proofs (e.g., Mejía-Ramos & Inglis, 2009) in individual or socially-embedded learning settings (e.g., Tristanti, Sutawidjaja, As'ari, & Muksar, 2016). When evaluating a constructed proof, representing the performance of an individual, the judgement may be based on criteria that capture the ideas necessary to establish the desired conclusion (Selden, A. & Selden, 2013b), the type of the presented arguments (empirical vs. deductive arguments) (e.g., Harel & Sowder, 1998), the soundness of the logical structure of the proof, and its correctness regarding the interpretation and use of formal symbols (e.g., Dubinsky & Yiparaki, 2000; Weber & Alcock, 2004).

The criteria that have to be taken into account when focusing on conjecturing and proving processes and that allow to infer whether an individual performs the processes that are required to engage successfully in a conjecturing and proving task have not been defined so far. Even though various models of the proof construction process have been developed in the past (which we will present in the next section), there is a lack of criteria that determine what successful conjecturing and proving processes are. Much of the extant work regarding conjecturing and proving processes involves theoretical arguments (e.g., Boero, 1999) and case studies (Zazkis et al., 2015). Systematic empirical investigations on the process characteristics that indicate how an individual can come to an interesting conjecture and an acceptable proof and that may form the foundation for defining criteria to which argumentative processes may be judged are rare. We contribute to this issue in the following way:

By introducing the concept of process characteristics, we aim to identify and empirically investigate the characteristics of conjecturing and proving processes that may predict the quality of the final product. We assume that the process characteristics contribute to the variance in undergraduate student's performance, besides their prior knowledge on proof.

8.4 Focusing on processes – Which characteristics are considered as relevant for successful conjecturing and proving processes?

Different frameworks and models have emerged describing the proof construction process and associated activities from an individual-mathematical perspective (cf. Boero, 1999; Carlson & Bloom, 2005; Schwarz et al., 2010). Just like other scientific disciplines, mathematics lives from the exchange of knowledge. External impulse and the discussion with peers present important opportunities to work on conjecturing and proving tasks, as the ideas of others may encourage to rethink already produced arguments or to develop new ones (Vogel et al., 2016). Therefore, the social-discursive dimension and related frameworks of collaborative argumentation need also to be taken into account when analysing argumentative activities (Osborne, 2010).

As stated, systematic empirical analyses about the effectiveness and success of specific activities within conjecturing and proving processes, in particular, in collaborative settings are rare. However, in the literature it has been argued that some process characteristics are relevant for high-quality conjecturing and proving performance or for collaboration. We will focus on these process characteristics in the next sections by presenting different frameworks and models from the individual-mathematical and social-discursive perspective.

8.4.1 The individual-mathematical perspective

The frameworks related to conjecturing and proving are largely based on theoretical assumptions or have been derived from the observations or self-reports of mathematicians (e.g., Schwarz et al., 2010) or students (e.g., Carlson & Bloom, 2005) on how they make sense of a problem, raise conjectures, and generate solutions. Though the models are illustrated in form of phases, they should not appear as linear. They can encompass cycles, refinements, impasses, or derivations (Furinghetti & Morselli, 2009).

The four phases of problem-solving

As conjecturing and proving are considered problem-solving processes, it is plausible to use the Multidimensional Problem-Solving Framework created by Carlson and Bloom (2005) which incorporates the four phases already outlined by Polya (1945): (I) understanding the problem, (II) developing a plan, (III) carrying out the plan, and (IV) looking back. Each of these phases involves cognitive and metacognitive activities and some of them bridge to the subsequent phase.

Phase I: The logical understanding of the problem is considered as a central characteristic of the first phase of problem-solving. Based on the observations of undergraduate students, Selden, J. et al. (2014) claimed that logical understanding may be reached by unpacking the conclusion, that is, looking up its definitions and adapting them to the given problem (Selden, J. et al., 2014). Reformulating the initial problem-situation by expressing the given task in one's own words using examples, symbols, pictures, or gestures is also regarded as a relevant approach in attempting to understand the problem (Furinghetti & Morselli, 2009).

Phase II: In the planning phase, the choice of appropriate representations and the ability to switch flexibly from one representation to another have been approved to be important activities to get from the initial hypothesis to the desired conclusion (Gholamazad, Liljedahl, & Zazkis, 2003). Boero (2001) argued that some representations may encourage the transformation of the given mathematical structure, whereas others may be obstructive to achieve the final envisaged shape.

Phase III: The study by Zazkis et al. (2015) has shown that the choice of the proving method and the amount of time that provers spend to explore each method could also have an impact on the success. Constructing new mathematical objects or manipulating variables without exactly knowing the value of these activities is sometimes necessary to generate ideas and to bring in new perspectives (Selden, J. et al., 2014).

Phase IV: Savic (2015b) used the framework of Carlson and Bloom (2005) to investigate the differences between the proving processes of mathematicians and that of graduate students.

One conclusion of his study was that checking and re-evaluating previous proving steps by critically analysing impasses and correcting one's own work are crucial proving activities. He proposed that the activities of questioning the constraints of conjectures, constructing counterexamples, and identifying gaps can be considered as influential factors regarding the success of proving processes.

Structural characteristics of arguments - Toulmin's model

In the literature, Toulmin's model (1958) has been applied both for investigating and for demonstrating how learning progresses in classrooms (Krummheuer 1995; Yackel 2001) and how argumentations develop (Wood, 1999). It divides the components of an argument into six elements: claims (assertions that have to be proven), data (facts that are used as evidence to support the claim), warrants (rules or statements that support the relationship between data and claim), qualifiers (restrictions under which the claim is assumed to stay true), backings (justifications for the warrant), and rebuttals (specific conditions in which the justification of the warrant would have to be repealed). This scheme is commonly used to analyse the structure of arguments (e.g., Clark, D. B. et al., 2007) and proofs from a cognitive point of view (Pedemonte, 2008).

The phases of argumentation and proof construction

Boero's model of argumentation and proof construction (1999) points to the continuity between conjecturing and proving (*Cognitive Unity of Theorems*: Garuti et al. (1996)) and includes phases of exploration and systematization.

Phases of exploration: The explorative phases are important to get in touch with the problem, to notice regularities, to create a conjecture, and to gain insights into the concepts and definitions involved (e.g., Pinto & Tall, 2002). Koedinger (1998) observed that students often do not have many problems in formulating conjectures in general, but in developing conjectures that say "something new, something beyond repeating the premise" (p. 329). A sophisticated strategy to find more interesting conjectures is to vary elements of successive examples systematically and to investigate which relationships appear invariant (Koedinger, 1998). Examples can be used in a number of ways within the proof construction process. Some successful students generate a set of examples in order to search for patterns and to explore the underlying mathematical structure. They often strive to discover structural regularities and transform them into formal representations (Ozgur et al., 2017). In contrast, less experienced students often do not look for links between concept definitions and the examples they have produced and consequently, have problems to formalize empirical arguments (Alcock & Weber, 2010).

Phases of systematization: Engaging in activities such as elaborating, syntactifying, and rewarranting have been put forward as promising strategies to build explicit links between informal and formal arguments. In this case, elaborating means adding more details to the proof, that has to be constructed, by making the warrants of informal arguments explicit and by enclosing supporting (empirical) data and warrants to them. Changing a graphical representation of a concept or argument to a verbal-symbolic one comprises the activity of syntactifying, and rewarranting describes the replacement of an informal warrant by a new, more suitable, verbal-symbolic warrant and thus, allows to make formal inferences (Zazkis et al., 2016). Other researchers also claimed that informal ideas and arguments have to be linked to formal definitions (e.g., Pinto & Tall, 2002) and that the rules of logic have to be applied (Selden, J. et al., 2014). Boero (1999) described the phases of systematization as processes of organizing single arguments into a deductive chain and of communicating them to the scientific community.

8.4.2 The social-discursive perspective

Although mathematical argumentation often occurs in social contexts, the models and frameworks presented above conceptualize problem-solving, argumentation, and proving as individual activities rather than as social practices. Considering argumentation as a social activity raises the question which activities characterize good collaboration and how collaboration influences students' individual conjecturing and proving processes.

Modes of collaboration

The term collaboration implies a joint production of a solution, where students share their ideas, build upon each other's arguments, and contrive a collective understanding of a problem (Staples, 2007). Several researchers agree that collaboration and social interactions are essential for mathematical problem-solving and especially for the development of students' argumentation skills (e.g., Alrø & Skovsmose, 2002; Vogel et al., 2016; Yackel & Cobb, 1996). Mueller et al. (2012) distinguished three different modes of collaboration. The first form describes the co-construction of arguments in which students alternately produce one joint argument. Integration, the second mode, occurs when learners take the ideas and contributions of others into account to strengthen their own initial arguments. The third one, modification, emerges when originally incorrect or unclear arguments of learners are corrected by their learning partners. In their study, Mueller et al. (2012) observed that all three modes influenced students' mathematical argumentation processes in a strong way.

Building a common ground

The success of any mode of collaboration might depend on different aspects such as actively building a common ground (Clark, 1996), sharing information, ideas, and cognitive processes (Hinsz, Tindale, & Vollrath, 1997), engaging in transactive discussions (Teasley, 1997; Vogel et al., 2016), or developing common plans of how to approach a specific task or goal (Barron, 2000). Meier et al. (2007) assumed that attempts to reach a mutual understanding, such as eliciting feedback from the learning partner or asking questions for clarification, can be considered as indicators for successful collaboration as well as coordinative aspects in order to avoid confusion about whose turn it is to bring in new ideas or to monitor the remaining time to solve the task. Collecting all task-relevant information that is spread over the participants and relating it to the concepts that already have been established during discourse are deemed to be important as well as motivational factors and the interpersonal relationship between learners.

Interactivity

A further characteristic that is often cited in the context of collaboration is the interaction-sequence of learners participating in an argumentative discourse. Leitão (2000) proposed that the pattern of generating an argument, constructing a counterargument, and creating a reply that builds upon the counterargument and modifies the initial argument shapes the process of social knowledge construction and the transformation through argumentation in a desirable way.

8.4.3 Summary

From the literature we have sourced, we conclude that exploring conjectures critically, generating accurate and structurally sound arguments as well as linking different ideas and arguments to each other and to relevant theoretical concepts are characteristics of conjecturing and proving that may have an impact on the success. Employing explicit feedback strategies, such as paraphrasing or ensuring mutual attention by asking questions as well as taking the learning partner's contributions into account and evaluating them critically, may lead to a deep elaboration of the problem situation and simultaneously may foster the development of evidence-based arguments. However, whether this is true is subject to empirical investigation.

We term the four individual-mathematical process characteristics *critical exploration, accuracy and precision, argument structure*, and *cognitive complexity*, the three social-discursive ones *critical questions, turn-taking-sequence*, and *reaction to the learning partner's errors* (see Study II- Chapter 7).

8.5 The current study

In this study, we analysed the interaction of prospective mathematics students working in dyads on a conjecturing and proving task from an individual-mathematical and social-discursive point of view. The primary aim of this study was to examine students' prior knowledge, their individual-mathematical and social-discursive process characteristics of collaborative conjecturing and proving (that are based on the research presented above), and their final products in attempt to better understand how successful conjecturing and proving proceed. This approach allows us to find empirical evidence to determine characteristics of successful conjecturing and proving processes. In addition, we are interested in investigating whether the process characteristics describing the collaborative partners' interactions correlate with each other.

In particular, we investigated the following research questions:

(RQ1) Which individual-mathematical and social-discursive process characteristics of collaborative conjecturing and proving go along with the success on the resulting conjecture and proof (the final product)?

We assume that overall individual-mathematical and social-discursive process characteristics predict overall quality of the final product (H1). From an individual-mathematical point of view and based on research on conjecturing and proving (e.g., Boero, 1999; Koedinger 1998, Furinghetti & Morselli 2009, Zazkis et al., 2016), we expect that *all four individual-mathematical process characteristics* work together (in response to changing demands during the collaborative conjecturing and proving process) to directly impact the final product (H1.1). From the social-discursive perspective and based on research on collaborative problem-solving (e.g., Blanton & Stylianou, 2014; Roschelle & Teasley, 1995; Vogel et al., 2016), we assume that the *three social-discursive process characteristics* mutually affect the quality of the final product as well (H1.2). Regarding the individual-mathematical process characteristics, we propose more differentiated hypotheses: We suspect that the *critical exploration* process characteristic influences the quality of the formulated conjecture (H1.1.1). For instance, varying examples systematically and testing the constraints of a conjecture have been found to positively influence its content and correctness (e.g., Koedinger, 1998; Leuders & Philipp, 2013; Pease, A. & Martin, 2012). In addition, we assume that the cognitive complexity process characteristic is an indicator of the number of generated ideas that are needed to solve the proving task (H1.1.2). This hypothesis is based on previous findings that reformulating the problem situation, creating analogies to other tasks, and linking ideas and arguments to each other and to formal definitions facilitate the development of different ideas and arguments related to the proving task (Alcock & Inglis, 2008; Furinghetti & Morselli, 2009; Pease, A. & Martin, 2012). Furthermore, we expect that the argument structure process characteristic predicts the overall quality of the final product and, particularly, the soundness of the structure of the final proof product (H1.1.3). Thereby, we follow the assumption that the generation of arguments based on facts and supported by warrants enables the construction of evidenced-based conjectures and proofs (e.g., Pedemeonte, 2008), a transition from informal to formal arguments (Zazkis et al., 2016), and thus, the construction of proofs that correspond to shared mathematical standards. Regarding the accuracy and precision process characteristic, we hypothesize that it influences the overall quality of the final product, especially the correctness of formal symbols used to present the proof (H1.1.4). It can be argued that applying definitions and rules in a precise way as well as checking and re-evaluating previous proving steps may help to draw correct inferences (Savic, 2015b; Selden, J. et al., 2014) and to present conjectures and proofs in an accurate manner (e.g., Weber & Alcock, 2004).

(RQ2) Do the relations between the process characteristics of collaborative conjecturing and proving and the final product remain substantial when prior knowledge on proof is controlled? Do the process characteristics mediate between prior knowledge of proof and the final product?

According to other researchers, we assume that the (learning-) prerequisite prior knowledge on proof may influence conjecturing and proving processes (e.g., Selden, A. et al., 2010) as well as the final product (e.g., Sommerhoff et al., 2016; Ufer et al., 2008; Weber, 2001) (H2.1). As the proposed set of process characteristics is not intended to be exhaustive, a direct and an indirect effect of prior knowledge on proof on the final product are expected (H2.2)

(RQ3) In collaborative settings, do the process characteristics among partners correlate to each other?

Collaborative settings provide opportunities to reflect and adapt one’s thinking (Cohen, 1994; Webb, Troper, & Fall, 1995) and to share mathematical knowledge and strategies in attempting to influence the thinking of others (e.g., Barron, 2000; Mueller et al., 2012; Yackel & Cobb, 1996). Martin and Towers (2003) emphasized that the emergent mathematical understanding should be understood as something more than merely an individual activity, namely as a phenomenon that grows through social interaction. From this perspective, it can be assumed that the individual-mathematical process characteristics describing the interactions of both learners correlate to each other (H3.1) as well as the individual-mathematical process characteristics of one learner and the social-discursive process characteristics of the collaborative partner (H3.2).

8.6 Method

8.6.1 Participants and Design

N = 98 prospective mathematics students (51 male, 3 not reported; M_{age}=19.83, SD_{age}=3.56) from two German universities took part in the present study. The instructional settings were embedded in a voluntary two-week preparatory course for university mathematics. The course was offered before the beginning of the first semester to support incoming students in the transition to university mathematics. It comprised twelve lectures and fourteen tutor exercises on basic topics such as propositional and predicate logic, proof techniques, elementary number theory, relations, and induction. The number of course attendees was much higher, but we excluded participants that missed the "prior knowledge on proof"- test or the collaborative learning session.

On the third day of the course, the students individually worked for 45 minutes on a self-designed test on their prior knowledge on proof. The collaborative learning session, which was recorded and finally analysed, took place on day six. In this session, students were assigned to homogenous dyads with respect to their prior school achievement and instructed to work collaboratively in a computer-supported-learning environment on an open-ended conjecturing and proving task. After around 50 minutes, at the end of the collaborative learning session, they were asked to write down their individual solutions.

8.6.2 Materials and Instruments

Prior knowledge on proof in elementary number theory

To assess prospective students' prior knowledge on proof, we adapted a test that was already used in several prior studies (e.g. Kollar et al., 2014; Schwaighofer et al., 2017). The test covered four facets of mathematical argumentation and proof skills, namely, five items measuring technical proof skills, four items testing flexible proof skills and eight items measuring conjecturing and proving skills. As technical proof skills we understand the application of simple rules, whereas flexible proof items require more than one proving step and additionally, the change of the representational system. The conjecturing and proving items of this test involve the evaluation of true and false mathematical statements and the subsequent construction of a proof or refutation. All of the items are elementary number theory problems (example items can be found in Figure 10).

5 items - technical proof skills (e.g. Show that for natural numbers a and b the following statement is true: If 15 divides (10a-5b) then 3 divides (2a-b))
4 items - flexible proof skills (e.g. Prove the following statement: The sum of a natural number and its square plus one is odd.)
8 items - conjecturing skills for false and true statements (e.g. Prove or refute the following statement: Sum up the square of an odd number and the square of an even number. If you divide this result by 4 the remainder is always 1.

Figure 10: Example items of the test for assessing students' prior knowledge on proof. The test was adapted from Reichersdorfer et al., 2012.

To score the quality of students' proof attempts, a four-level coding was applied. Missing or irrelevant trials were scored with zero. Partially correct answers that included less than half of all central arguments required to completely solve the task were coded with one, and solutions that consisted of more than half of all central arguments but with small methodological errors (like an incorrect proof structure) received score two. Completely correct argumentations and proofs were scored with score three. Two independent raters coded all items and the inter-rater reliability for each part was found to be good ($M_{\text{K, PKP}}$ = .86, $SD_{\text{K, PKP}}$ =.08).

Collaborative mathematical and conjecturing processes

During the collaborative learning session, the participants were seated in dyads working on the following conjecturing and proving task: *"Take four consecutive numbers, multiply them, and add one. Repeat this and try to find similarities. Formulate a conjecture and prove it!".* The task was presented in a computer-supported learning environment that allowed the students to write down their individual ideas and arguments, to exchange them in a shared workspace, and to verbally discuss their approaches. The whole working process was recorded and later analysed by two independent raters. For the analysis a high-inference coding scheme was developed that allowed to assess the process quality along four individual-mathematical (*accuracy and precision, argument structure*, *cognitive complexity, critical exploration*) and three social-discursive process characteristics (*critical questions, turn-taking-sequence, reaction to the learning partner's errors*) (see Study II – Chapter 7). Since quality assessments require some interpretation by the rater and thus, often result in a low reliability, a detailed coding handbook including descriptions about the different quality levels of each process characteristic was written and used in an extensive rater training. In this way we tried to standardize judgements as much as possiblTa to reach a high inter-rater-reliability. Each scale of a process characteristic comprised five different quality levels that went from 1 (excellent) to 5 (very poor). The raters were allowed to take notes of their impressions while watching the recordings to make it easier to remember and to evaluate the learner's interaction. In order to reduce the memory load, we split each video into two equal-length parts that were rated separately for each learner of one dyad. A short summary of the rating scales of the process characteristics can be found in Table 11. The interactions of twenty undergraduate students were coded by two independent raters. As a measure of inter-rater reliability, we calculated the intra-class-correlation for each process characteristic. All ICCs were above 0.7 and thus, allow a meaningful interpretation (Wirtz & Caspar, 2002). We calculated the overall mean of the four individual-mathematical process characteristics ($M_{\text{individual-mathematical}} = 2.849$, $SD_{\text{individual-mathematical}} = .417$) and the overall mean of the three social-discursive process characteristics ($M_{\text{social-discursive}} = 2.938$, $SD_{\text{social-discursive}} = .340$). Both overall mean values indicated a satisfactory quality level and correlated significantly ($r = .412$; $p < .001$). The standard deviations gave some evidence that we were able to measure the variance within the interactions of each learner during their collaborative discourse. The reliability of the overall individual-mathematical scale was $\alpha = .73$ and the reliability of the overall social-discursive scale was $\alpha = .51$.

The quality of the final product

We used the individual solutions that each student created after the collaborative learning session to measure the students' conjecturing and proving performance and analysed them with regard to the quality of the formulated conjecture and the quality of the constructed proof. The quality of the conjecture was assessed along one criterion reaching from 0 (incorrect) to 2 (correct and creative). For evaluating the quality of the proof, we used three different criteria. The first one captured the number of arguments produced (0: less than half of all arguments required to completely solve the task, 1: more than half of all arguments, but not all arguments, 2: proofs that include all relevant arguments). The quality of the structure of the proof was measured along the second proof criterion (0: proofs that showed large structural gaps, 1: proof with small structural gaps, 2: proofs that were valid from a structural point of view). The third proof criterion assessed the quality of the use of mathematical symbols and language (0: large formal errors, 1: small formal errors that have no influence of the meaning, 1: formally correct arguments). After the coding process the four criteria were aggregated to a product score that served as the measurement-value for the quality of the final product (presenting the conjecturing and proving performance of a student). The reliability of the overall product scale was $\alpha = .73$. Double coding of over 10% of the data led to the inter-rater reliability of $M_{\kappa,\text{ product}} = .78$ ($SD_{\kappa,\text{ product}} = .13$).

Table 11: *Seven process characteristics of collaborative conjecturing and proving processes and the resulting scales of the rating scheme.*

	Description of the scale
Content accuracy	Focuses on the extent to which mathematical arguments and ideas are captured correctly and accurately, and on the extent to which a learner re-evaluates previous proving steps by critically analyzing impasses and mistakes.
Argument structure	Measures whether mathematical arguments are specified by basing claims on facts, making warrants explicit, and using qualifiers to point out under which conditions the argument is assumed to stay true.
Cognitive complexity	Focuses on how key arguments are developed and combined at different phases of the proof construction process. Thereby, activities such as changing the representation system or cross-referencing within different ideas play an important role.
Critical exploration	Refers to the extensive investigation of conjectures by generating examples and counter-examples, exploring their limits, and formulating more than one initial conjecture.
Critical questions	Refers to all contributions that critically question the ideas and arguments of the learning partner or one's own proving steps and that demand explicit feedback that is attempted to be implemented.
Turn-taking-sequence	Measures whether the learner takes the partner's contributions into account either by extending or refining the learning partner's ideas or by strengthening his / her own arguments.
Reaction to the learning partner's errors	Focuses on the extent to which the learner addresses the errors and impasses produced by the learning partner and measures whether the learner finds explanations for these errors or proposes alternative solution steps.

8.7 Results

8.7.1 Relations between process characteristics and the quality of the final product

To investigate the impact of different process characteristics of collaborative conjecturing and proving on the quality of the resulting product, linear regression analyses were conducted (RQ1) using M*plus* Version 7.4 (Muthén & Muthén, 2015) and taking the hierarchical structure of the data (students nested in dyads) into account. We assumed that the overall mean value of the individual-mathematical process characteristics as well as the overall mean value of the social-discursive process characteristics predict the quality of the resulting conjecture and proof. Therefore, the product score that assessed the quality of the resulting conjecture and proof served as dependent variable in each of the analyses. Results showed that the overall mean value of the individual-mathematical process characteristics had a significant impact on the overall quality of the resulting conjecture and proof (supporting our assumption H1.1;

$\hat{\beta}$ = -.392; p < .001). This could not be hold for the social-discursive component of argumentation. Here, the overall mean value of the social-discursive process characteristics was no predictor for the overall quality of the final product (refuting our assumption H1.2; $\hat{\beta}$ = -.043; p = .832) (see Table 12).

Since these results did not completely meet our expectations, we proceeded with some further explorative analyses. We investigated whether there is a significant interaction between individual-mathematical and social-discursive process characteristics on the quality of the final product. Therefore, we conducted a multiple linear regression analysis with the overall mean value of the individual-mathematical process characteristics, the overall mean value of the social-discursive process characteristics, and their respective product term as predictive variables and the product score as dependent variable. To facilitate the interpretation of the results, we centred the predictive variables by subtracting the sample means of the individual-mathematical and social-discursive components from their respective values. Again, the overall mean value of the individual-cognitive process characteristics showed a significant impact on the overall quality of the final product ($\hat{\beta}$ = -.349; p < .001), but no significant effect of the social-discursive component could be found ($\hat{\beta}$ = -.005; p = .972). The interaction between individual-mathematical and social-discursive process characteristics was statistically significant ($\hat{\beta}$= .320; p = .003) indicating that students with higher social-discursive process quality profited more from the positive impact of the individual-mathematical process characteristics on the overall quality of the final product than students with lower social-discursive process quality.

To examine the hypotheses (H1.1.1) and (H1.1.2) we conducted two single probit regression analyses. A significant impact of critical exploration on the quality of the produced conjecture could be discovered (supporting our assumption H1.1.1; $\hat{\beta}$ = -.321; p = .034) as well as a significant impact of cognitive complexity on the number of arguments (supporting our assumption H1.1.2; $\hat{\beta}$ = -.370; p = .002) (see Table 12).

Regarding the two content-specific process characteristics (argument structure; accuracy and precision), our assumptions H1.1.3 and H1.1.4 could be supported. Results revealed that the process characteristic accuracy and precision had a significant influence on the quality of the resulting product ($\hat{\beta}$ = -.449; p < .001) and in particular on the soundness of the structure of the final proof (supporting our assumption H1.1.3; $\hat{\beta}$ = -.471; p < .001). As expected, the accuracy and precision process characteristic predict the overall quality of the final product ($\hat{\beta}$ = -.369; p < .001, especially the correctness of formal symbols used to present the proof (supporting our assumption H1.1.4; $\hat{\beta}$ = -.299; p = .007).

Furthermore, we observed that the process characteristics critical exploration and cognitive complexity were no predictors for the overall quality of the final proof ($\hat{\beta}$ = -.111; *p* = .391; $\hat{\beta}$ = -.187; p = .090).

8.7.2 Relations between process characteristics, the quality of the final product, and prior knowledge on proof

Regarding the second research question (RQ2), we examined whether the relation between individual-mathematical process characteristics and the final product remain substantial if prior knowledge on proof is controlled. We investigated the direct effect of prior knowledge on proof on the quality of the final product and the indirect effect mediated by individual-mathematical process characteristics. The quality of the final product served as the measurement value of students' conjecturing and proving performance.

Descriptive results of students' prior knowledge on proof and the final product can be found in Table 13 and 14. The participants seemed to have less problems to identify and refute false statements than to solve technical proof items, flexible proof items or to evaluate and prove true statements. The overall mean value of prior knowledge on proof was M_{PKP} = 1.414 (SD_{PKP} = .587). This implies that on average less than half of all arguments required to completely solve the items were present.

A path analysis was conducted to estimate the influence of prior knowledge on proof on the individual-mathematical process characteristics and the final product (representing students' conjecturing and proving performance). Prior knowledge on proof had a significant impact on the overall mean value of the individual-mathematical process characteristics (supporting our assumption H2.1; $\hat{\beta}$ = -.5188, $\hat{z}$ = 4.999, *p* < .001). Moreover, a statistically significant direct effect ($\hat{\beta}$ = .238, $\hat{z}$ = 2.036, *p* = .042) and an indirect effect linking prior knowledge on proof and the final product by the overall mean value of the individual-mathematical process characteristics ($\hat{\beta}$ = .139, $\hat{z}$ = 2.112, *p* = .035) were found, confirming our assumptions (H2.2). Overall, the specified model explained 19.5% of the variance of conjecturing and proving performance, and 26.8% of the variance of the overall mean value of the individual-mathematical process characteristics.

8.7.3 Relations between the process characteristics of students within one dyad

To analyse the correlative relationship between the individual-mathematical process characteristics of both students, and the relationship between the individual-mathematical process characteristics of a student and the social-discursive process characteristics of the collaborative partner (RQ3), we restructured the dataset and considered both halves of the working process separately. In the first half, no significant relationship between the individual-mathematical process characteristics of a student and the social-discursive process characteristics of the collaborative partner could be observed (r = .106; p = .398). Also, the individual-mathematical process characteristics of both learners seemed to be unrelated. Yet, in the second half of the working process, results indicated that the individual-mathematical process characteristics of the learner and the social-discursive process characteristics of the respective learning partner were significantly related to each other (r = .433; p < .001). Additionally, we found a highly significant relationship between the individual-mathematical process characteristics of both students (r = .648; p < .001) at this part of the working process.

Table 12: *Results of linear regression analyses predicting the quality of the resulting product.*

Depended variable	Predictor	β	$S.E._{\beta}$	$\hat{\beta}$	p
Overall quality product	$M_{ind\text{-}math}$	-.316	.068	-.392	<.001
	R^2	.103			
Overall quality product	$M_{soc\text{-}dis}$	-.039	.185	-.043	.832
Overall quality product	$M_{ind\text{-}math}$	-.278	.113	-.349	.013
	$M_{soc\text{-}dis}$	-.005	.134	-.005	.972
	$M_{ind\text{-}math} \times M_{soc\text{-}dis}$	.335	.085	.320	<.001
	R^2	.248			
Overall quality product	*accuracy and precision*	-.226	.053	-.369	<.001
	R^2	.137			
Overall quality product	*argument-structure*	-.259	.054	-.449	<.001
	R^2	.202			
Overall quality product	*cognitive complexity*	-.101	.057	-.187	.078
Overall quality product	*critical exploration*	-.001	.071	-.111	.392

Product criterion: conjecture	*critical exploration*	-.331	.151	-.321	.103
	R^2	.103			
Product criterion: number of arguments	*cognitive complexity*	-.379	.116	-.370	.001
	R^2	.137			
Product criterion: structure	*argument structure*	-.257	.061	-.471	<.001
	R^2	.222			
Product criterion: correctness	*accuracy precision*	-.172	.065	-.299	.007
	R^2	.089			

Table 13: *Mean values and standard derivations for the four quality facets of prior knowledge on proof.*

	M (SD)
Technical proof skills	1.381 (.740)
Flexible proof skills	1.259 (.807)
Conjecturing true statements	1.355 (.715)
Conjecturing false statements	1.660 (.787)

Note: min=0; max=3

Table 14: *Mean values and standard derivations for the quality criteria of the resulting proof.*

	M (SD)
Quality of the conjecture	1.337 (.592)
Number of arguments	1.211 (.845)
Structure of the proof	1.158 (.621)
Correctness of mathematical symbols	1.084 (.804)

Note: min.=0; max=2

8.8 Discussion

The present study examined the interactions of prospective undergraduate mathematics students working collaboratively on a conjecturing and proving task and sheds light onto the relations between different individual-mathematical and social-discursive process characteristics, the quality of the resulting product, and the prior knowledge on proof. Even though the concepts of argumentation and proof (e.g., Balacheff, 1999; Boero et al., 1996; Mariotti, 2006) as well as students' difficulties within conjecturing and proving (e.g., Martin & Harel, 1989; Moore 1994; Epp, 2003) have already been objects of extensive studies, this article addressed a largely unexplored field in mathematics educational research, namely the empirical investigation of different process characteristics of collaborative conjecturing and proving and their interplay in predicting conjecturing and proving performance, respectively the quality of the resulting product.

Most studies on conjecturing and proving were based on small sample sizes and qualitative findings (e.g., Koedinger, 1998; Savic, 2015b; Selden, A. et al., 2010) that did not allow any statistical analyses. Therefore, we saw the need for systematically examining the impact of process characteristics that are assumed to be important from a theoretical point of view on the quality of the resulting product (by controlling for prior knowledge on proof) in order to identify key characteristics for successful collaborative conjecturing and proving processes. Furthermore, we contribute to the mathematics educational literature by studying individual-mathematical and social-discursive process characteristics and by analysing the relations between these process characteristics of one student and the process characteristics of the collaborative partner.

8.8.1 The identification of key characteristics of successful collaborative conjecturing and proving processes

We extracted different individual-mathematical and social-discursive process characteristics of collaborative conjecturing and proving from the literature that were assumed to play a crucial role for the success. Our results indicate that the process characteristics related to the individual-mathematical component of mathematical argumentation predict the quality of the resulting product. This coincides with our hypothesis (H1.1) that grounded on prior research (e.g., Pease, A. & Martin, 2012; Savic, 2015b; Selden, A. & Selden, 2013b). For instance, Savic (2015b) has already claimed that “the proof-construction process could have a lasting effect on the correctness of a student's proofs” (p. 81). In addition, we found empirical evidence for our assumptions that ideas and arguments have to be formulated with precision (H1.1.3) and based on facts and warrants (H1.1.4) during the collaborative discourse in order to succeed. These results supported the claims and findings of other researches that evaluating the truth or falsity of mathematical statements requires the correct use of definitions, concepts, and rules (e.g., Epp, 2003; Moore, 1994) as well as the ability to express mathematical ideas in a largely formal correct way (e.g., Selden & Selden, 2011). In addition, it supported the finding that highly-successful students are able to identify and rethink incorrect assertions or impasses (Weber, 2009). Moreover, prior studies have already shown that the construction of explicit warrants (Zazkis et al., 2016) and the additional use of modal qualifiers to justify deductive and non-deductive conclusions are crucial for the process of solving proof problems (Inglis et al., 2007). Furthermore, our study points out that the critical exploration of conjectures by generating (counter-) examples and testing their limitations is a predictive criterion for the quality of the produced conjecture (H1.1.1), confirming the observations by other researchers (Koedinger, 1998; Pease & Martin, 2012). Regarding the process characteristic cognitive complexity, we found empirical support for our assumption that changing the representation system, creating analogies to other tasks, and linking ideas and arguments to each other and to formal concepts can encourage students to bring in new perspectives in the proof construction process and thus, to generate arguments that are needed to solve the proving task (H1.1.2). Other researchers also emphasized the importance of working flexibly with different representations (e.g., Alcok & Inglis, 2008) and of combining different problem-solving steps (e.g., Hiebert et al., 2003) to produce arguments.

Even though the three social-discursive process characteristics (critical questions, turn-taking-sequence, reaction to the learning partner’s errors) addressed aspects that were supposed to be relevant for collaboration quality from a theoretical point of view (e.g., Meier et al., 2007; Roschelle & Teasley, 1995) and that in turn may positively affect students’ conjecturing and proving performance (cf. Mueller et al., 2015), no significant impact on the final product could be observed. On the one hand, this refutes the assumption (H1.2) that students who build upon each other's arguments and who question the ideas of their learning partners show a better performance at the end (cf. Kneser & Ploetzner, 2001). However, some researchers of other disciplines also were not able to find substantial correlations between collaborative process quality and solution quality (Meier et al., 2007; Sampson & Clark, 2009). One reason might be that some learners do not refer to each other's contributions at all, but still construct knowledge (constructive activities) (Chi, 2009; Chi & Wylie, 2014). Another explanation might be that some learners who are initially interested in the collaborative partners' work, but do not receive explanations from them that are understandable and expedient, start to "self-regulate their learning behaviour by making more use of their internal and less use of their social resources" (Kneser & Ploetzner, 2001, p. 79) after some time.

On the other hand, the interaction effect of individual-mathematical and social-discursive process characteristics indicates that students who show a high individual-mathematical process quality and high collaboration quality (above the overall average values respectively) may benefit more from their individual-mathematical activities than students who are able to present the same individual-mathematical process quality but lower collaboration quality. It seems that the impact of high-quality individual-mathematical process characteristics on the quality of the resulting product might be only shown in combination with high collaboration quality. The process of explaining and questioning ideas may stimulate learners to re-evaluate their own arguments and thus, might even make themselves aware of possible impasses in their reasoning (van Boxtel, van der Linden, & Kanselaar, 2000). Productive collaboration between successful students may also sustain the motivation to explore conjectures more deeply and to rethink already produced arguments (Roschelle & Teasley, 1995). Moreover, it is assumed that "embedding a conjecture in a different body of knowledge can lead to further insights into the conjecture" (Pease, A., Smaill, Colton, & Lee, 2009, p. 133).

In summary, we can say that we have identified different process characteristics that predict the quality (or at least one quality criterion) of the resulting product. Especially the individual-mathematical process characteristics can be considered as key characteristics of successful collaborative conjecturing and proving processes. Regarding the social-discursive process characteristics, it can be deduced that collaboration is not necessarily sufficient for the initial performance on an argumentation task (Meier, et al., 2007; Sampson & Clark, 2009), but that the impact of collaboration process quality may also depend on the individual-mathematical process characteristics.

8.8.2 Mediation effects of individual-mathematical process characteristics of collaborative conjecturing and proving processes

One further purpose of this study was to investigate whether collaborative conjecturing and proving processes link prior knowledge on proof to conjecturing and proving performance. The results of this study show that collaborative conjecturing and proving processes that can be assessed by different process characteristics mediate between prior knowledge on proof on the one hand and the final product on the other hand (H2), and thus support the framework proposed by Blömeke et al. (2015). This framework suggests that processes are mediating elements between cognitive (and affect-motivational) dispositions and the situation-specific performance. Prior knowledge on proof as (learning-) prerequisite represented the cognitive dispositions related to mathematical argumentation skills, the final product represented the performance in a collaborative conjecturing and proving situation. We examined this mediating mechanism as well as the direct effect of prior knowledge on proof on conjecturing and proving performance. Our findings provided empirical evidence for all of these theoretical considerations and revealed the overall mean value of the individual-mathematical process characteristics as a significant mediator. The estimated direct effect of prior knowledge on proof on the quality of the resulting product confirmed that individual-mathematical process characteristics do not completely mediate the impact of prior knowledge on proof on conjecturing and proving performance. This result indicates that conjecturing and proving processes as “sequence[s] of mental and physical actions” (Selden, A. et al., 2010, p. 205) may contain crucial processes that are not directly accessible (e.g., Abrahamson & Lindgren, 2014), but that may explain something beyond the performance and thus, it might be possible to infer “kinds of difficulties in students’ proof construction processes from their written proof attempts” (Selden, J. et al., 2014, p. 246). This has to be examined in further studies. Overall, we can conclude from the path analysis that prior knowledge on proof predict processes and performance in specific situations (cf. Spencer & Spencer, 2008) and that individual-mathematical process characteristics play a mediating role within this relationship.

8.8.3 Relations between dyad partners

It is generally accepted that collaborative work between peers provides a promising environment for problem-solving and learning (e.g., Kaartinen & Kumpulainen, 2002; Kneser & Ploetzner, 2001; Roschelle & Teasley, 1995). In this study we were interested in analysing whether the individual-mathematical contributions of a learner and the social-discursive contributions of the collaborative partners are related to each other. Our results suggest that the relationship between the process characteristic of learners working collaboratively on a conjecturing and proving task changes during the discourse. The contributions of both learners seem to be more affected by their learning prerequisites at the beginning, but their individual-mathematical process characteristics became more interrelated over the time as well as the social-discursive contributions of one learner and the individual-mathematical contributions of the corresponding partner. One reason could be that students initially have to find out which role they want to assume during the discourse and which knowledge they already possess or need to solve the given task (Kneser & Ploetzner, 2001). After this initial phase, it has been shown that the learners align themselves in their cognitive contributions and that there is a positive relationship between the quality of social-discursive activities of one learner and the individual-mathematical process characteristics of the dyad partner. These findings are in accordance with the observations of Anjewierden et al. (2011) that students within a collaborative learning setting “mirror” each other in such a way that if one learner contributes a more domain-related message, the other learner also tends to post a domain-related statement. We concur with other researchers that collaboration could influence the building of (mathematical) arguments and that also the way how students interact during the discourse depend on their prior knowledge and process characteristics (Kneser & Ploetzner, 2001; Mueller et al., 2012).

8.8.4 Restrictions and perspectives

This exploratory study provides empirical evidence for the effectiveness of different process characteristics that have been assumed to be relevant for successful conjecturing and proving processes from a theoretical point of view. We tested theoretical assumptions about individual-mathematical and social-discursive process characteristics that were primarily based on findings of prior case studies. Even though our study broadens the literature on conjecturing and proving by systematically examining these process characteristics and by presenting findings that go beyond small-sample-sized, qualitative research studies and purely theoretical considerations, there are some limitations that have to be mentioned. By analysing the interactions of undergraduate students working collaboratively on one conjecturing and proving task, we can only describe which process characteristics within this single context have an effect on the quality of the resulting product. The contextualized character of argumentation skills, which means that person- and situation-specific factors, and especially the interplay between these both factors, play an important role (Koeppen et al., 2008), could not be taken into account in this study. To manage this problem, several tasks should be used to investigate students’ conjecturing and proving processes in future studies. A first step might be to increase the number of tasks in the context of number theory and then to present conjecturing and proving task from other context areas such as geometry (e.g. Komatsu, 2011). Modifying the context with regard to the intended proving activities may allow to detect differences between the (learning-) prerequisites and processes that are required to solve proof-reading, proof-validation or proof-construction tasks (cf. Mejía-Ramos & Inglis, 2009; Selden, A. & Selden, 2015; Sommerhoff, 2017). A further restriction is that our study is confined to individual-mathematical and social-discursive aspects, neglecting affective-motivational dispositions and processes (Koeppen et al., 2008; Weinert, 2001). It might be fruitful to extend our coding-scheme by operationalizing affective-motivational process characteristics and by taking meta-cognitive processes, such as reflection, (Kneser & Ploetzner, 2001) into account.

With respect to our analyses, a major limitation lies in the small sample size and in the relatively small number of process characteristics. SEM analyses with latent variables, which require certainly more participants than the equivalent model with one observed variable and a representing number of individual-mathematical and social-discursive process characteristics, would be preferable, since single indicators (respectively process characteristics) are usually measured with errors (Wolf, Harrington, Clark, & Miller, 2013). Additionally, due to the small sample size, the results of the moderation analysis and correlations analysis (with the reconstructed dataset) have to be interpreted carefully.

Finally, regarding our social-discursive process-characteristics, we did not check how exactly students build upon the ideas of their learning partners. This means that we did not differentiate between *co-construction*, *integration* and *modification* (Mueller et al., 2012) or between *dialogic* and *dialectic dialogues* (Vogel et al., 2016; Wegerif, 2008). It would be interesting if a more detailed analysis of the social-discursive component would present the assumed relevance of collaboration quality better.

8.8.5 Conclusions

Conjecturing and proving processes are considered as core components in which mathematical argumentation skills become (at least partially) accessible (cf. Blömeke, 2015) and which substantially affect the quality of the resulting product (representing conjecturing and proving performance) (e.g., Savic, 2015b). Knowledge of the dependencies of these processes on (learning-) prerequisites and their impact on the quality of the final conjecture and proof appears crucial for teaching and interventions in schools and higher mathematics education to foster the development of mathematical argumentation skills. As a starting point we suggested and empirically investigated a set of individual-mathematical and social-discursive process characteristics that have been assumed to describe successful collaborative conjecturing and proving processes from a theoretical point of view. From our study's findings, we draw the following conclusions:

1) Individual-mathematical process characteristics of collaborative conjecturing and proving have a significant impact on the quality of the resulting product.

Generating structurally sound arguments that are based on facts, supported by warrants that strengthen the connection of the facts to the claim, and that are classified by qualifiers (e.g., Clark, D. B. et al., 2007; Inglis et al., 2007; Toulmin, 1958) turned out to be key characteristics of successful conjecturing and proving processes as well as the accurate and precious formulation of mathematical ideas and arguments (e.g., Reinholz, 2016; Savic, 2015b) during the collaborative discourse. Exploring conjectures critically (e.g., Koedinger, 1998; Pease, A. & Martin, 2012) and linking different ideas to each other and to formal concepts (e.g., Pease, A. & Martin, 2012) appeared also to be essential, but particularly for a specific criterion representing the quality of the produced solution.

2) Social-discursive process characteristics show no significant impact on the quality of the resulting product, yet the interaction of individual-mathematical and social-discursive process characteristics does.

Against the theoretical assumptions, the social-discursive process characteristics have not directly influenced the quality of the resulting conjecture and proof. However, our results point to the significant effect of the interaction of individual-mathematical and social-discursive processes, indicating that collaboration quality is necessary, but not sufficient for the success of collaborative conjecturing and proving processes.

3) Individual-mathematical process characteristics (partially) mediate the impact of prior knowledge on proof on the quality of the final product.

From this finding, we conclude that conjecturing and proving processes (beyond the outcome) could provide additional information about students' mathematical argumentation skills, but that the inclusion of processes involves further problems as mental processes sometimes remain hidden (e.g., Blömeke et al., 2015; Selden, A. et al., 2010).

4) The relations between the process characteristics of a dyad group become stronger during the argumentative discourse.

The results of this study indicate that differences of how students within one dyad engage in collaborative conjecturing and proving processes and utilize their prior knowledge become less important during their collaborative work as they adjust their process characteristics to each other during the argumentative discourse (cf. Anjewierden et al., 2011).

8.8.6 Theoretical and practical implications

In view of our results, at least in the context of the study, specific focus might be given to the individual-mathematical process characteristics of collaborative conjecturing and proving, in particular on generating accurate and structural sound arguments during the argumentative discourse. It might be useful making students aware of these activities by explicitly discussing them in (advanced) undergraduate mathematics courses and to design interventions that address these process characteristics. For instance, it would be a good starting point to request students to construct warrants paired with appropriate modal qualifiers (Inglis et al., 2007) for relatively easy conclusions or to ask them to analyse the structure of a written proof. Instructors can explicitly demonstrate their ways of generating logically sound arguments in their lectures and point out why the warrant they have constructed is needed (Harel & Sowder, 1998). Regarding the correct and accurate formulation of arguments, we suggest that students should discover and discuss different interpretations of definitions and arguments (Selden, J. et al., 2014), and that instructors explicitly convey how arguments can be accurately formulated by using, for instance, formal notations.

As we have extracted and identified some key characteristics of successful conjecturing and proving processes, it would be possible to adapt our coding-scheme as a diagnostic instrument which teachers and tutors may use to detect students' problems within conjecturing and proving and to find out where support is needed. This would allow a real-time analyses of students' (collaborative) conjecturing and proving processes. In addition, learners may be asked to evaluate their own collaborative conjecturing and proving processes along these key characteristics of collaborative conjecturing and proving in order to scaffold learning and encourage meta-cognitive skills (Lee, E. Y. C., Chan, & van Aalst, 2006; Meier et al., 2007).

Based on our findings regarding the interaction effect of individual-mathematical and social-discursive process characteristics, it would be promising to promote social-discursive process characteristics in addition to the individual-mathematical ones (cf. Vogel et al., 2017) and to design scaffolds that address both components of collaborative conjecturing and proving processes. As the social-discursive process characteristics did not directly affect the quality of the final product, researchers may investigate whether collaborative learning settings would be more beneficial for the long-term learning of mathematical argumentation and proof than for the instant performance on a conjecturing and proving task (cf. Sampson & Clark, 2009).

From our analysis of the mediation effect of prior knowledge on proof on the resulting product by individual-mathematical process characteristics, we deduce that it would be possible to scaffold the (learning-) prerequisites that are necessary to effectively engage in collaborative conjecturing and proving processes in order to stimulate the individual-mathematical process characteristics.

Considering the relations of process characteristics between dyad students, it would be beneficial to deliberately choose the "tuning-in-phases" for interventions, as these phases may provide opportunities for better and more innovative argumentation processes. However, whether the students jointly converge to a higher or lower quality level remains a question for further investigations.

9 Discussion

9.1 Summary of the central findings

Much research on mathematical argumentation and proof has focused on the products, the accuracy and validity of the conjectures and proofs themselves (e.g., VanSpronsen, 2008). While this is a valuable endeavour, it is just as important to identify and investigate the characteristics of successful conjecturing and proving processes.

In this dissertation, we presented three studies, one literature review and two empirical analyses, on (collaborative) conjecturing and proving processes. We will summarize the theoretical framing and central findings of the three studies and, where possible, show links between them.

9.1.1 Theoretical framing

The main goal of this dissertation was to systematically describe and analyse process characteristics of collaborative conjecturing and proving and to investigate their relevance for the resulting product (the quality of the formulated conjecture and constructed proof). The first chapter provided a short summary of previous research' findings about students' difficulties with conjecturing and proving. In the second chapter, an individual-mathematical and social-discursive perspective on mathematical argumentation and proof were presented. We compared different definitions of both constructs and drew attention to the three elements (resources, processes, and situations respectively products) of argumentation skills. Conjecturing and proving processes were conceptualized as specific types of argumentation processes, consisting of an individual-mathematical and social-discursive component (cf. Kollar et al., 2014). The individual-mathematical component captured phases of exploring the problem situation, formulating a conjecture, selecting appropriate arguments and organizing them into a deductive chain (e.g., Boero, 1999; Reichersdorfer et al., 2012). Phases of communicating and presenting one's own arguments (e.g., Hanna, 1990) as well as evaluating and integrating the arguments of others (e.g., Chi, 2009; Leitão, 2000) were allocated to the social-discursive component of conjecturing and proving processes. We set out an analytic framework for measuring mathematical argumentation skills by differentiating between (learning-) prerequisites, processes, and the final product. In particular, we were interested in figuring out how conjecturing and proving processes can be assessed. We suggested to focus on process characteristics that have been mentioned in the literature on (mathematical) argumentation, taking both components of conjecturing and proving into account, and investigating their connection to the quality of the final proof products. Conjecturing and proving process characteristics were considered as observable variables that mediate between prior knowledge on proof and the final product (including the formulated conjecture and the constructed proof).

9.1.2 Central findings of the literature review

Our research review provided an elaborated summary of common claims and empirical findings of previous research regarding high-quality conjecturing and proving processes. It started with the application of a topic-modeling method, resulting in a clustering of different research topics on conjecturing and proving and the extent to which each of these topics has been researched in the field of mathematics education. The algorithm suggested to differentiate 17 topics associated with the concepts of conjecturing and proving. The most common topics are “the nature of proof and teaching of proof” topic, followed by the “university level” topic and “the school students and teachers” topic. According to the three perspectives on proving outlined by Stylianides, G. J. et al. (2017) namely the problem-solving perspective, the proving as convincing perspective, and the proving as a socially-embedded activity perspective, we attempted to allocate each topic to one of these research perspective on conjecturing and proving. We observed that the studies that are most representative for one topic typically share one common perspective on conjecturing and proving. This means that they are based on similar theoretical constructs, investigate comparable research questions, and sometimes even apply similar research methods. Besides the three established perspectives on proving, we identified a new one, namely the discovery perspective. This perspective fits with the discovery function of proof, as described by Villiers (1999). Furthermore, our findings support the observation of Stylianides, G. J. et al. (2017) that some articles and research reports cannot clearly be allocated to one single perspective on conjecturing and proving. This especially concerns the “nature of proof and teaching of proof” topic or, for instance, the “argument structure” topic. The facts that the “nature of proof and teaching of proof” topic itself discusses the role and functions of proof and that most articles and research reports are composed of words from this topic may explain why it covers different perspectives on conjecturing and proving. Articles and research reports that represent the “argument structure” topic focussed on the cognitive and structural continuities and distances between argumentation and proof and hence pointed to a topic that can be elucidated from all perspectives. Overall, the topic model provides a first systematization of the large body of research on conjecturing and proving. Subsequently, the claims and findings of the most representative articles and research reports for each topic (in total 45 papers) have been analysed qualitatively and synthesized to explore which processes related to conjecturing and proving are considered crucial for the success. We decided to distinguish between the sub-goals within conjecturing and proving processes that are often described in more general terms, difficult to operationalize, and the process characteristics of conjecturing and proving that are assumed to be helpful in accomplishing these sub-goals. Several studies, independently of their research topics, demonstrate consistency in the sub-goals (related to conjecturing and proving) which they considered as necessary intermediate steps within conjecturing and proving processes. We extracted eleven different types of sub-gaols from the literature on conjecturing and proving, incorporating some sub-goals related to broader categories (for instance, the sub-goal “inventing and formulating new conjectures or refining existing conjectures” can be assigned to an overarching category “exploring the problem-situation”) and some sub-goals that define a new category such as the sub-goal “recovering from impasses/ avoiding errors”. Regarding the process-characteristics of conjecturing and proving, we have identified nine categories that capture those characteristics that have been described as observable processes and as processes that occur in the context of successfully achieving one or more of these sub-goals. For instance, the process characteristics category “example use” that includes process characteristics such as “varying examples systematically”, “considering boundary cases”, “choosing examples with specific properties”, “testing a diversity of examples”, and “attempting to construct counterexamples” was frequently discussed in the context of the sub-goal “developing a strong understanding of the proof/ estimation of the truth” or of the sub-goal “intervening and formulating new conjectures or refining existing conjectures”, but also in the context of the sub-goal “drawing inferences” or the sub-goal “communicating and presenting arguments” (cf. Ellis et al., 2017; Ozgur et al., 2017). Process characteristics such as “splitting a task into more, simpler sub-tasks” (Pastre, 1989) or “working forwards from the givens and working backwards from what one is required to show” (cf. Carroll, 1977; Küchemann & Hoyles, 2006) form the process characteristics category “domain-general problem-solving processes that, for instance, occur in the context of the sub-goal “structuring and organizing” or “recovering form impasses/ avoiding errors”. In this way, we categorized the previous studies’ claims and findings about promising conjecturing and proving processes with regard to sub-goals and process characteristics of conjecturing and proving. The framework we created addresses both dimensions, the sub-goals and process characteristics, and therefore demonstrates the range of processes that may occur during conjecturing and proving, and how they may be employed in a deliberate and productive manner.

9.1.3 Summary of the technical report

In addition to the theoretically comprehensive overview of the sub-goals and process characteristics of conjecturing and proving that our literature review provides, we have set ourselves the aim of empirically investigating various types of process characteristics. This implies to overcome some methodological challenges, such as the development of a rating scheme that allows to quantify undergraduate students’ collaborative conjecturing and proving processes. We focussed on undergraduate students' collaborative conjecturing and proving processes for the following reasons: (i) there is a lot of research on undergraduate students' (un-) successful proof attempts, what facilitates the extraction and defining of theory-based process characteristics; (ii) collaborative learning situations allow to measure processes (verbal and written utterances) directly at the time and place of their natural occurrences (without explicitly prompting students to verbalize their thoughts).

The *Technical Report* describes a new rating scheme to assess undergraduate mathematical students' collaborative conjecturing and proving processes empirically and the methods used to apply this assessment tool. It shows how guidelines for high inference ratings from the educational science have been adapted to capture process characteristics of mathematical argumentation empirically. The report includes information on the rater training, documentations of the ratings scales, and describes the coding procedures in detail.

9.1.4 Central findings of the empirical study

As we lack quantitative studies concerning mathematics students' activities in generating mathematical conjectures and proofs that focussed on the processes and the resulting products, we conducted an exploratory empirical study. In our empirical study, we have operationalized and analysed seven process characteristics of collaborative conjecturing and proving. These process characteristics comprised the *critical exploration* of the problem situation, the *cognitive complexity* of students' contributions, the *accuracy and precision* of how arguments are presented, the *argument structure*, the *critical questioning* of the arguments produced by the learning partner, the *turn-taking structure* within the collaboration process, and the *reaction to the learning partners' incorrect or invalid arguments*. The process characteristics related to the first four categories were inferred from the mathematics educational literature that primarily conceptualizes conjecturing and proving as individual-mathematical activities (e.g., Koedinger, 1998; Selden, J. et al., 2014; Zazkis et al., 2015), the latter three predominately from the research of the Learning Sciences and psychology with a social-discursive perspective on argumentation (e.g., Asterhan & Schwarz, 2007; Chi & Wylie, 2014; Weinberger & Fischer, 2006). Based on these two perspectives on conjecturing and proving, we were interested in empirically studying the relationship between individual-mathematical and social-discursive process characteristics of collaborative conjecturing and proving. Furthermore, we focussed on their impact on the quality of the resulting product (the formulated conjecture and the constructed proof). We investigated *N=98* prospective undergraduate mathematics students' collaborative conjecturing and proving processes, their prior knowledge on proof, and their final proof products resulting from these processes.

Empirical study – Study II

We draw *four central new findings* from our fist analyses:

The results show that (1) individual-mathematical and social-discursive process characteristics can be reliably measured, indicating that the developed rating scheme allows the user to assess the quality of collaborative conjecturing and proving processes from an individual-mathematical and social-discursive perspective on mathematical argumentation. In addition, this study demonstrates that (2) the two components of collaborative conjecturing and proving processes can be empirically distinguished from one another and that they were not strongly related. The findings of Kollar et al. (2014) also indicted that undergraduate students' performance on the individual-mathematical respectively social-discursive component of mathematical argumentation skills are largely unrelated to each other. There is, to our knowledge, no further prior research on the structure of collaborative conjecturing and proving processes that would allow a comparison of our results. However, the independence of both components might have been strengthened by the differences in their conceptualizations, as the individual-mathematical and social-discursive process characteristics were derived from distinct research disciplines, from the mathematics educational respectively psychological and the Learning Sciences discipline. The disciplines may differ in terms of the conceptualizations and operationalisations they use (cf. Inglis & Foster, 2018). Nevertheless, we have tried to find a common language to transfer the identified process characteristics into rating scales. In addition, this study makes also a methodological contribution. Taking into account that the participants of our study worked in dyads on a conjecturing and proving task, the analysis of the within-cluster and between-cluster variance, induced by the hierarchical structure in the data, shows that (3) the clustering of the data has to be corrected by using multi-level analysis (in particular, concerning the process data that emerged from the second part of the collaboration process). Yet, results also indicate that (4) it is necessary to consider each student's contributions separately (cf. Anjewierden et al., 2011), and not only aggregated data on dyad level, at least for the process characteristics related to the categories cognitive complexity, critical questions and turn-taking-sequence. To sum up, studying the empirical structure of individual-mathematical and social-discursive process characteristics of collaborative conjecturing and proving is completely innovative as no prior comparable research on mathematical argumentation ca be found.

Empirical study – Study III

We draw *eight central new findings* from our second analyses:

This analysis has examined the theoretically assumed effects of the individual-mathematical and social-discursive process characteristics on the quality of the resulting product.

Furthermore, we were interested in investigating the effect of prior knowledge on proof on the final product and the extent to which it is mediated by the conjecturing and proving processes that we operationalized by the set of process characteristics. Our hypothesis that all theory-based process characteristics related to the seven categories are predictive for the quality of the resulting product was only partly confirmed. The empirical findings have shown that (1) prospective undergraduate mathematics students' individual-mathematical process characteristics can be used to predict their performance in solving conjecturing and proving tasks, (2) but not the social-discursive process characteristics. The latter may be due to the fact that in the studies from which they were discussed previously, the socio-discursive process characteristics were originally conceptualized as collaborative learning activities rather than as activities that directly promote success in achievement situations (e.g., Chi & Wylie, 2014; Vogel et al., 2016). However, we found (3) an interaction effect of individual-mathematical and social-discursive process characteristics, indicating that a good collaboration is necessary, but not sufficient for the success. Regarding the individual-mathematical process characteristics of collaborative conjecturing and proving, the findings of this study go in line with the results of our research review. In particular, we found that (4) generating accurate and structurally sound arguments during the proving discourse has a substantial impact on the quality of the resulting product, thus confirming hypotheses derived from prior qualitative studies (e.g., Inglis et al. 2007; Krummheuer, 1995; Pedemonte, 2008; Yackel, 2002). (5) Process characteristics related to the critical exploration of the problem situation have been found to be relevant for the formulation of correct and creative conjectures, process characteristics related to the cognitive complexity category for the number of relevant mathematical ideas visible in students' solutions. These results provide empirical evidence for the respective relationships between process characteristics and sub-goals, as described in our research review, and therefore support the claims and findings of other researchers (e.g., Ellis et al., 2017; Kidron & Dreyfus, 2014; Koedinger, 1998). Supporting our basic assumption that process characteristics are strongly influenced by situational instantiations of prior knowledge on proof, we observed (6) strong correlations between students' individual-mathematical process characteristics and their prior knowledge on proof. Furthermore, we found empirical evidence for the assumption that (7) conjecturing and proving processes mediate between the prior knowledge on proof (representing the students' underlying resources) on the one hand and the quality of the resulting product (representing the students' performance in generating conjectures and constructing proofs) on the other hand (cf. Blömeke et al., 2015; Sommerhoff, 2017), and therefore support for our analytical framework. A further result of this study was that in the second half of the collaboration process, the individual-mathematical process characteristics of both students who worked together within a dyad were significantly correlated, suggesting firstly the relevance of taking the hierarchical structure of the data (students nested in dyads) into account (as already mentioned in the context of study II). Moreover, this finding indicates that the students of a dyad align their cognitive behaviour during the collaboration discourse (e.g., Anjewierden et al., 2011). Additionally, we observed (8) a highly significant relationship between the individual-mathematical process characteristics of one student and the social-discursive process characteristics of his or her learning partner. This result shows that the students' individual-mathematical process characteristics should not be interpreted in isolation, in particular at the second half of the collaboration process. It provides evidence for the assumption that the individual-mathematical process characteristics could not be attributed only to the student who enacted them. The individual-mathematical process characteristics may need to be assessed with regard to the socio-discursive process characteristics of the learning partner, and more generally, in terms of the community in which the student participates (cf. Stahl, 2010).

9.2 Limitations, strengths, and implications for teaching and future research

This section aims to highlight and discuss the central findings of this thesis with regard to their strengths and limitations. We will put the results in a larger context of current research and draw some theoretical and practical implications for future research on conjecturing and proving.

9.2.1 Structuring the literature on conjecturing and proving

Our research review provides a new systematization of the literature on conjecturing and proving by presenting the main topics that have been investigated in the context of conjecturing and proving and discussing the research perspectives from which conjecturing and proving processes have been conceptualized. In consistency with the categorization suggested by Stylianides, G. J. et al. (2017), we identified several studies that extended on the problem-solving literature from cognitive psychology and conceptualized conjecturing and proving as problem-solving, several studies that built upon the literature on proof schemes and conceptualized conjecturing and proving as convincing, and a small proportion of studies that form the literature on conjecturing and proving as social practices and thus conceptualized conjecturing and proving as socially-embedded activities. Furthermore, we detected some studies that were based on the literature that described conjecturing and proving as tools to generate and explore new knowledge and that conceptualized conjecturing and proving as discovering. Even though the statistical clustering of the studies to topics is quite objective, as the topic modeling analysis itself is conducted algorithmically, the interpretations of the identified topics are more subjective and therefore open to criticism (e.g., Inglis & Foster, 2018). One may also criticise that the allocation to the research perspectives rest on subjective judgements and that they should be verified by other researchers. However, the topic modeling method and the subsequent allocation to the topics and research perspective allow to uncover the topics and research perspectives on conjecturing and proving that appear to be under-researched so far. Based on those findings, we see the need for expanding the literature on conjecturing and proving that addresses the questions of how conjecture generation and proof construction activities are practiced in mathematical and classroom communities.

The higher-order sub-goals categories, which we have systematized with regard to the intermediate steps within conjecturing and proving processes deemed to be necessary for the successful generation of conjectures and proofs in the primary studies, provide a further "data-driven" structuring of the literature on conjecturing and proving. This structuring depends on the reviewer's insights and judgments as well. Yet, it may allow other researchers to save a lot of time and effort by using this categorisation in order to select the literature that is specifically designed to learn more about a particular sub-goal.

9.2.2 Bringing different research disciplines and perspectives together

In this thesis, we integrated knowledge from different research disciplines. Scientific communities from different research disciplines usually use distinct theoretical descriptions of similar phenomena and apply distinct research methods (e.g., Inglis & Foster, 2018). In attempting to find out what *good* collaborative conjecturing and proving processes are and how they can be defined, we reviewed the mathematics educational literatures as well as the psychological and the Learning Sciences literature on (mathematical) argumentation. Even though we have identified different research perspectives on conjecturing and proving in the mathematics educational literature (e.g. the problem-solving perspective or the conjecturing and proving as discovering perspective), the processes associated with conjecturing and proving within this discipline were conceived primarily as activities to be performed by individuals (cf. Balacheff, 1988). The educational-psychological and the Learning Sciences research on scientific argumentation has its research focus predominately on the externalised processes and products of scientific argumentation and reasoning within social contexts (Fischer et al., 2014). Therefore, the process characteristics from an individual-mathematical perspective on conjecturing and proving (the *individual-mathematical process characteristics*) were primarily derived from the mathematics educational literature on conjecturing and proving, the social-discursive ones from the research in psychology and the Learning Sciences on scientific argumentation. Considering the findings from the different research disciplines allowed us to connect und contrast several individual-mathematical and social-discursive process characteristics of collaborative conjecturing and proving. However, one may argue against our approach on the grounds that each discipline differs in the way of how they use and operationalize terms and concepts. We are aware of the fact that those differences have to be taken into account. Yet, we have followed the recommendation of other researchers that more interdisciplinary research on (mathematical) argumentation is needed (e.g., Fischer et al., 2014; Sommerhoff, 2017) and that researchers should “make a concerted effort to read journals from outside the discipline” (Inglis & Foster, 2018, p. 495). In this way, we have combined the previously unconnected research traditions and thus contribute to a more comprehensive understanding of the (relatively similar or at least related) phenomena.

Furthermore, we successfully adapted the guidelines and procedures proposed by researchers from educational science (e.g., Gartmeier et al., 2014; Seidel, 2005) and psychology (Meier et al., 2007) to develop a high-inference rating scheme to assess the quality of collaborative conjecturing and proving processes. To our knowledge, high-inference rating schemes or comparable, more generic assessment methods have rarely been developed or used in the study of students’ mathematical cognitions, so far.

9.2.3 Proposing the “sub-goals versus process characteristics” framework

In our literature review, we present a framework that shows how different processes, and in particular the co-occurrence of these processes, can contribute to the successful generation of conjectures and proofs. Since simply applying a certain kind of proving process type does not necessarily lead to success (e.g., Furinghetti & Morselli, 2009; Zazkis et al., 2016), we have tried to uncover the process characteristics of conjecturing and proving that are assumed or reported as being helpful in achieving the inter-mediate steps within conjecturing and proving processes (the “sub-goals”) that, in turn, are considered as being necessary for the successful generation of conjectures and proofs. We categorized and synthesized claims and findings of prior research on conjecturing and proving with regard to these two dimensions (the sub-goals dimension and process characteristics dimensions). The resulting framework demonstrates the process characteristics and sub-goals that these researchers indicate are used by successful provers. We claim that our “sub-goals versus process characteristics” framework may be adapted for teaching purposes. Teachers and lecturers may explain the identified sub-goals and highlight the ways in which students can succeed when employing the associated process characteristics of conjecturing and proving.

Moreover, the new framework may be used to guide research on conjecturing and proving as it provides a structure for conceptualizing and designing future research studies. The framework offers a theoretical foundation for addressing rich and detailed questions about students’ conjecturing and proving processes. For instance, it provides a conceptual structure for empirically investigating which characteristics of proof processes go along with acceptable proof constructions, which type of example use goes along with the development of understanding of the problem situation, or which problem-solving strategies go along with resolving fixations. In addition, the framework demonstrated current research foci regarding process characteristics and sub-goals and how these are examined and related to each other in the mathematics educational research literature. Some combinations of process characteristics and sub-goals were not addressed at all. This could either be an indication that the combinations are still unexplored or under-researched, or that the process characteristics are not expected to encourage the achievement of these sub-goals. Moreover, this framework provides an initial overview about which process characteristics, according to the current literature, might be encouraged in order to help students to fulfil a specific sub-goal within conjecturing and proving processes and thus constitutes a theoretical base for developing scaffolds and interventions.

Besides the strengths of this framework, one has to be aware that it was necessary to go beyond the contents of the original studies at some points (for instance, when the content of the study not directly focussed on sub-goals or process characteristics of conjecturing and proving) in order to produce a satisfactory synthesis (cf. Marston & King, 2006).

9.2.4 The analytical framework of assessing argumentation skills

As our research review has shown that quantitative studies on (collaborative) conjecturing and proving are rare, we designed one empirical study including two main analyses. Our central aims were to examine a set of process characteristics of collaborative conjecturing and proving that were suggested by prior research to be predictive for students' performance in solving conjecturing and proving tasks. To address these aims, we adapted the framework for mathematical argumentation and proof skills proposed by Sommerhoff (2017). This framework distinguishes the resources (prerequisites) that underlie the argumentation and proof skills, the processes by which the skills become enacted in corresponding situations, as well as students' performance in these situations. The situation we chose was one conjecturing and proving task that the students had to solve collaboratively initially, but for which they had to compose an individual solution afterwards. We followed the definition of argumentation skills suggested by Kollar et al. (2014) and expanded the framework by distinguishing between individual-mathematical and social-discursive processes of collaborative conjecturing and proving. Both, the individual-mathematical and social-discursive processes were considered as mediating elements between the resources and the performance in this specific situation. Another change we made was that we additionally expected a direct effect of the prerequisites on students' performance, since we could not ensure to capture all characteristics of argumentation processes in our analysis, which are relevant for the quality of the final conjecture and proof (performance).

We measured the performance by requesting the students to write down an individual solution of the conjecturing and proving task the students had collaboratively worked on before. Each solution was evaluated with regard to the quality of the formulated conjecture and the quality of the constructed proof. To examine the influence of individual-mathematical and social-discursive processes on the quality of the resulting product, we operationalized a set of theory-based process characteristics of collaborative conjecturing and proving. Regarding the resources, we adapted the argumentation and proof test develop by Reichersdorfer et al. (2012) that consisted of seventeen items representing students' prior knowledge on proof.

Even though we investigated all three components and their relations in one of our empirical studies, our analytical and practical implementation of the framework for mathematical argumentation and proof skills has some limitations. First, we considered only one single situation by using one task. Second, we measured the resources by a test that did not allow a finer-grained distinction of the students' cognitive resources. Previous research has shown that it is important to discern between the types of resources predictive for students' performance (e.g., Chinnappan et al., 2012; Sommerhoff et al., 2016; Ufer et al., 2008), and that different situations may cause different processes (e.g., Mejía-Ramos & Inglis, 2009) and require different resources (e.g., Sommerhoff et al., 2017). Hence, future studies should try to focus on the resources more precisely and to assess the processes related to proof-construction, proof-reading, and proof presentation. The relations between specific resources (such as methodological knowledge) and process characteristics (such as the process characteristic argument structure) have not been investigated in detail

However, to our knowledge, this dissertation provides the first study that investigates the relevance of process characteristics in a systematic way, at least for one concrete situation. Furthermore, it demonstrates how the three components of argumentation skills can be operationalized and how their relationships can be empirically investigated.

9.2.5 Developing and evaluating an instrument for assessing the quality of collaborative conjecturing and proving processes

The analysis of conjecturing and proving processes as well as of collaboration is a central topic in mathematics educational (e.g., Mueller et al., 2015; Weber, 2004) respectively psychological or the Learning Sciences research (e.g., Schwaighofer et al., 2017; Vogel et al., 2016). However, defining and operationalizing process characteristics that constitute key characteristics of successful conjecturing and proving and developing instruments capable of assessing these characteristics is a challenging task (cf. Meier et al., 2007). In this dissertation, seven process characteristics extracted from the literature were defined, operationalized, and rated quantitatively: *critical exploration, cognitive complexity, accuracy and precision, argument structure, critical questioning, turn-taking-sequence, reaction to the learning partner's mistakes.* In extracting and defining these process characteristics, we followed a top-down approach. We set what we wanted to observe on the basis of theoretical assumptions and empirical finings of previous studies and created detailed descriptions of the process characteristics. Based on these theoretical assumptions, we developed a rating scale that distinguishes five quality levels for each process characteristic. In doing so, we followed the guidelines of developing high-inference rating schemes proposed by Seidel (2005). The endpoints of the rating scales were labelled as “very good” on their positive sides and as “very bad” on their negative sides.

The rating scheme was applied to process data of undergraduate students working collaboratively on a conjecturing and proving task. Study II and Study III, which investigated the empirical structure of the process characteristics and their relationships to prior knowledge on proof and to the quality of the resulting product, can be considered as evaluation studies for the newly-developed coding scheme.

Regarding the top-down approach to identify and define process characteristics of collaborative conjecturing and proving, one critical remark may be that it would be possible that we have overlooked some aspects that have made the given situation (in which our studies have been embedded, namely a dyadic setting where students were requested to collaboratively solve a conjecturing and proving task) specifically successful or unsuccessful. This might be a limitation of the coding scheme, but we aimed to define process characteristics that will probably be easily to transfer to other contexts. We argue that the process characteristics extracted from the literature will allow to compare a wider range of collaborative conjecturing and proving situations than process characteristics that would have been grounded in data gathered with one specific setting and task (cf. Meier et al., 2007).

Furthermore, we are aware of the fact that details of the collaborative conjecturing and proving processes may be lost due to the aggregation of multiple processes necessary for the application of high-inferential rating scales. More fine-grained, low-inference category systems would allow to detect the details, yet possibly at the expense of capturing the overall complexity of these processes (e.g., Newble, 2004; Seidel, 2005). However, since our goal was to assess and quantify the quality of collaborative conjecturing and proving processes on a relatively global level and not to calculate frequencies, a high-inferential rating scheme constituted the most suitable method (cf. Seidel, 2005; Wiesbeck, 2015).

Based on the positive findings regarding the inter-rater reliability, consistency, and the relations between the operationalized process characteristics and the quality of the resulting product (as a hint for the process-outcome validity), it would be possible to use our new developed rating scheme in different areas of collaborative conjecturing and proving.

This rating scheme allows judging the observed interactions of (undergraduate) mathematics students working collaboratively on conjecturing and proving tasks against defined quality levels, and thus can be applied in future studies to yield a direct evaluation of the quality of collaborative conjecturing and proving processes.

9.2.6 The importance to differentiate between individual-mathematical and social-discursive process characteristics

Along with Kollar et al. (2014), we differentiated between an individual-mathematical and social-discursive component of collaborative conjecturing and proving processes. To test our assumption that these components comprise differentiable conjecturing and proving processes, we defined four individual-mathematical and three social-discursive process characteristics representing the two dimensions. Confirmatory factor analyses based on the ratings were conducted to check whether the data fit the theoretical model. The result of these analyses supported our assumption that the two dimensions can empirically distinguished from one another and that they are weakly correlated to each other.

Our second confirmatory factor analysis, in which the process characteristic *reaction to the learning partner's errors* was treated as an individual-mathematical characteristic, demonstrated better fit indices than our initial model, in which this process characteristic was treated as a social-discursive one. This observation indicates that there might be some processes that reflect both individual-mathematical and social-discursive skills, and thus are at the interface of the two components of collaborative conjecturing and proving processes (cf. Kollar et al., 2014). Future studies may include more of these characteristics that involve both components of mathematical argumentation skills and investigate whether these characteristics form a continuum between individual-mathematical and social-discursive process characteristics of collaborative conjecturing and proving processes.

Furthermore, our findings have pointed out that the individual-mathematical process characteristics strongly depend on prior knowledge on proof. It can be expected that social-discursive process characteristics also depend on prior knowledge on - what Kollar et al. (2014) call - the social-discursive component of argumentation skills. This comprises a domain-overarching ability to construct arguments, counterarguments, and syntheses of different arguments. In the study by Kollar et al. (2014) the relationship between students' prior knowledge on proof (representing the individual-mathematical component of argumentation skills) and their knowledge on the sequence of an argumentation process (representing the social-discursive component of argumentation skills) has been investigated. Results of this study demonstrated that these prior knowledge components are only weakly correlated. This might explain why the process characteristics also show low correlations.

Regarding the effects of process characteristics on the quality of the resulting product, the findings of our study indicate again that it is advisable to distinguish between individual-mathematical and social-discursive process characteristics:

9.2.6.1 Empirical evidence for the relevance of individual-mathematical process characteristics

As the success of collaborative conjecturing and proving processes is determined by their outcomes, we investigated the relationships between the process characteristics and the quality of the final product (while controlling for prior knowledge on proof). Results indicated that the individual-mathematical process characteristics are predictive for the overall quality of the resulting product. In particular, generating accurate and structurally sound arguments appear to be key characteristics of successful collaborative conjecturing and proving processes.

Based on these findings, it would be possible to develop pedagogical concepts and interventions to support students' conjecturing and proving processes by, for instance, explicitly addressing these process characteristics in lectures (cf. Selden & Selden, 2013), by prompting students to base their claims on warrant and to use qualifiers (cf. Inglis et al., 2007), or by encouraging students to adjust the wordings to present arguments in an accurate way (cf. Savic, 2015b).

Moreover, we observed that the relationship between prior knowledge on proof and the final product is only partially mediated by the individual-mathematical process characteristics. This can be considered as a hint that our set of process characteristics is not exhaustive or that some processes are even not accessible for external observation (e.g., Blömeke et al., 2015; Selden, A. et al., 2010). However, it should be emphasized that our set of individual-mathematical process characteristics can explain some part of the relationship between students' prior knowledge on proof and the quality of the final product.

9.2.6.2 The (lacking) influence of social-discursive process characteristics

The social-discursive process characteristics did not turn out to be predictive for the quality of the final product. One reason might be that, in previous studies from which they have been deduced, the social-discursive process characteristics were mainly conceptualized as collaborative learning activities that have an impact on the disposition to use argumentation skills (e.g., Vogel et al., 2016) or to generate knowledge in learning situations (Weinberger & Fischer, 2006), but not as activities that are directly related to the success in performance situations. In our research review, we also allocated the social-discursive process characteristics to the sub-goal of achieve a common understanding. From this perspective, it might play a role that the collaborative working session in this study, from which the data were extracted, may have been perceived by students as a performance situation rather than a learning situation (cf. Schulmeiß, Seidel, & Meyer). However, whether the students experienced the collaborative working session more as a performance- than as a learning situation and whether this affects the impact of social-discursive process characteristics on the quality of the final product remains an open question. Future studies may also investigate whether the impact of social-discursive process characteristics would increase when students work together for a longer period of time (Staples, 2007).

Based on the finding of our study that the interaction between individual-mathematical and social-discursive process characteristics has shown a statistically significant effect on the quality of the final product, it would be effective to study interventions that scaffold the social-discursive process characteristics in addition to the individual-mathematical ones.

While we do not claim that the seven process characteristics exhaust the fully spectrum of necessary or sufficient characteristics for successful mathematical argumentation and proof, we believe this is a promising start that could be helpful both to university teachers (including lectures and teachers) as well as to researchers. In our literature review, we have identified a larger number of process characteristics that are described as (potential) predictors for the successful outcome from a theoretical point of view. Several of these process characteristics have not been empirically investigated, yet. These process characteristics can be addressed in future studies. Our rating scheme can serve as a template how these ‘potential predictors’ can be operationalized and assessed. Moreover, the process characteristics we have defined and operationalized can help tutors or lecturers to identify the specific challenges their students face and to find out where support is needed.

9.2.7 Individual learner vs. group level

The second part of the empirical study supports the assumption that relationships between the single students’ contributions to a dyadic proving process can be found (cf. Asterhan & Schwarz, 2009; Wecker & Fischer, 2011). However, a replication respectively analogous study with more participants would also allow to test whether students’ individual learning outcomes are predicted by the social-discursive and individual-mathematical process characteristics of the learning partner (as hypothesized, e.g., by Wecker & Fischer, 2011; Vogel et al., 2016).

When analysing students working in dyads on complex tasks, the question arises whether the students contribute equally to the (successful) outcome and whether the group-processes adequately represent the processes of the individuals. Based on our findings, we see the need to consider the individual’s contributions separately, and not on the dyad level. This go in line with the observations of other researchers that the quantity and quality of contributions that are made in a small group within a collaborative working process are not necessarily equally distributed among the members of the group (Jiménez-Aleixandre, Bugallo Rodríguez, & Duschl, 2000) and that the results from group-level analysis and individual-level analysis can differ (MacKinnon, Fairchild, & Fritz, 2007). However, regarding the within-cluster and between-cluster variance, our study indicates that the hierarchical structure of the data (students nested into dyads) has to be taken into account when conducting statistical analyses. Using methods to control nested data dependencies has proven to be a promising approach to analyse collaborative conjecturing and proving processes.

10 Appendix

10.1 Hoch-inferente Ratings - kooperative Argumentations- und Beweisprozesse

Allgemeine Anmerkung:

- Die Ratingskalen dienen dazu, kooperative Argumentations- und Beweisprozesse anhand ausgewählter theorie-basierter Prozessmerkmale zu bewerten.
- Die einzelnen Ratingskalen und deren Qualitätsniveaus werden anhand von Kriterien und Beispielen erläutert.
- Für jede Ratingskala gilt: Es müssen immer alle Aspekte der Qualitätsstufe, die in der Beschreibung angeführt werden, erfüllt sein, um den entsprechenden Code zu vergeben. Wenn diese „Mindestanforderung“ nicht erfüllt ist, muss ein Code für eine der unteren Stufen vergeben werden.
- Die Rater beobachten die Videoaufzeichnung bis zur Hälfte des Kooperationsprozesses und bewerten die Prozesse eines Lerners anhand der sieben Prozessmerkmale. Danach wird der zweite Teil angesehen und die Prozesse desselben Lerners bewertet. Die Rater können Teile der Videoaufnahmen nach Bedarf anhalten und beliebig oft ansehen. Für die Beurteilung der Prozesse des Lernpartners wird das Video erneut bis zu ersten Hälfte sowie bis zur zweiten Hälften angesehen und jeweils bewertet.

Fachliche Korrektheit der Beiträge zum Diskurs

Frage: Inwiefern bringt der Lerner fachlich korrekte Äußerungen in den Diskurs ein? Inwiefern werden also fachlich korrekte Aussagen über die betrachteten mathematischen Konzepte gemacht?

Kriterien: Bewertet werden die vom Lerner im Diskurs geäußerten Beiträge. Wesentlich ist dabei, ob es sich aus mathematischer Sicht – unter Berücksichtigung der üblichen Standards in einem Brückenkurs – um fachlich korrekte Äußerungen handelt. Es wird berücksichtigt, ob und wie schnell Fehler korrigiert werden und inwiefern der Problemlöseprozess durch die fachlich falschen Äußerungen des Lerners beeinträchtigt wird (d.h. inwiefern führen die fachlich falschen Äußerungen des Lerners zu längeren Fehlwegen).

1	Fachliche Beiträge des Lerners sind *weitgehend korrekt*, vereinzelte falsche Beiträge werden *umgehend selbst korrigiert*.
2	Fachliche Beiträge des Lerners sind weitgehend korrekt, vereinzelte falsche Beiträge werden *nicht immer umgehend selbst korrigiert*. Diese betreffen jedoch nur einzelne Stellen des Diskurses und *führen nicht zu längeren Fehlwegen* im Problemlöseprozess.
3	Der Lerner äußert *wenige falsche Beiträge*, die er nicht immer umgehend selbst korrigiert. Über einzelne Stellen hinaus führen diese *vereinzelt zu längeren Fehlwegen* im Problemlöseprozess, die aber im Wesentlichen selbst oder vom Partner erkannt und geklärt werden.

4	Der Lerner äußert *falsche Beiträge*, die er nicht immer umgehend selbst korrigiert. Über einzelne Stellen hinaus führen diese *mehrfach zu längeren Fehlwegen* im Problemlöseprozess, die nicht immer im Wesentlichen erkannt und geklärt werden.
5	Der Lerner äußert *viele falsche Beiträge*, die *überwiegend nicht korrigiert* werden. Es treten *mehrfach längere Fehlwege* im Problemlöseprozess auf, die im Wesentlichen nicht erkannt und geklärt werden.

Typische Beispiele:

Falsch: „Null ist doch keine gerade Zahl“

Bemerkungen: Die Korrektheit einer fachlichen Äußerung ergibt sich immer aus dem Kontext. Nicht-zielführende Äußerungen können somit dennoch fachlich korrekt sein.

Anmerkungen während der Kodiererschulung:

Anmerkung zu Code 1: Werden lediglich kleine formale Mängel, die die Bedeutung einer Aussage nicht wesentlich einschränken, nicht selbst korrigiert, so kann dennoch der Code 1 vergeben werden.

Anmerkung zu Code 3: Bsp.: Lerner verrechnet sich und findet den Fehler nicht. Dies führt zu einem längeren Fehlweg im Problemlöseprozess, der nicht mehr korrigiert wird. Alle weiteren Aussagen sind allerdings weitgehend korrekt -> Code 3

Allgemein: Notationsfehler werden ebenfalls berücksichtigt. Je nach Art des Fehlers und dessen Konsequenzen auf den Problemlöseprozess werden (Notations-)Fehler allerdings unterschiedlich gewichtet (d.h. inwiefern führt der Fehler zu einem falschen „Fehlweg“).

Kognitives Niveau der Beiträge zum Diskurs

Frage: Inwiefern bringt der Lerner zu unterschiedlichen Phasen des Argumentationsprozesses neue inhaltliche Ideen ein, stellt logische Beziehungen zwischen unterschiedlichen Ideen und Konzepten her und bringt somit neue Aspekte in den Diskurs ein?

Kriterien: Die Kodierung fokussiert insbesondere auf Stellen an denen Impasses auftreten oder an denen der Problemlöseprozess die Chance hat eine neue Richtung zu bekommen, weil neue Ideen eingebracht oder vorhandene Ideen und Konzepte neu verknüpft werden. Ideen bezeichnen hier im Diskurs vorhandene Aussagen, Zusammenhänge und Strukturierungen, aber auch andere inhaltliche Informationen aus dem Kontext des Brückenkurses bzw. dem individuellen Vorwissen der Lernenden.

Relevant ist dabei nicht, ob die eingebrachten Ideen tragfähig sind oder ob sie weiterhin aufgegriffen werden. Zentral ist das Einbringen von neuen Perspektiven durch die Verknüpfung von Ideen und Konzepten, die allerdings auf den Inhalt des Diskurses bezogen sein müssen (Problemstellung, Vorlesung, Tutorien, Vorwissen,...).

1	Der Lerner bringt *zu unterschiedlichen Phasen* des Argumentationsprozesses häufig *neue inhaltliche Ideen* ein, *verknüpft an den meisten geeigneten Stellen* vorhandene Ideen und Konzepte auf eine neue Art oder assoziiert neue inhaltsbezogene Ideen außerhalb des Diskurses, die er wiederum mit bereits vorhandenen Ideen und Konzepten in Verbindung bringt. *An vielen Stellen* bringt er damit *völlig neue Perspektiven* auf die gerade im Fokus stehenden Konzepte ein, so dass neue Beweisideen generiert werden können.
2	Der Lerner bringt zu unterschiedlichen Phase des Argumentationsprozesses häufig neue inhaltliche Ideen ein, verknüpft an den meisten geeigneten Stellen vorhandene Ideen und Konzepte auf eine neue Art oder assoziiert neue inhaltsbezogene Ideen außerhalb des Diskurses. *Selten, aber an mindestens an einer Stelle* bringt er damit *völlig neue Perspektiven* auf die gerade im Fokus stehenden Konzepte ein, so dass neue Beweisideen generiert werden können.
3	Der Lerner bringt *nur zu bestimmten Phasen* des Argumentationsprozesses *neue inhaltliche Ideen ein*, *verknüpft teilweise* an geeigneten Stellen vorhandene Ideen und Konzepte auf eine neue Art oder assoziiert neue inhaltsbezogene Ideen außerhalb des Diskurses. Allerdings bringt er damit im Wesentlichen *keine völlig neue Perspektive* auf die gerade im Fokus stehenden Konzepte ein, so dass keine neuen Beweisideen generiert werden können.
4	Der Lerner bringt nur zu bestimmten Phasen des Argumentationsprozesses neue inhaltliche Ideen ein, *verknüpft nur vereinzelt* an geeigneten Stellen vorhandene Ideen und Konzepte auf eine neue Art oder assoziiert neue inhaltsbezogene Ideen außerhalb des Diskurses. *Weitgehend wiederholt oder paraphrasiert* er bereits eingebrachten Ideen oder bereits bekannte Konzepte.
5	Der Lerner bringt *nur vereinzelt neue inhaltliche Ideen* ein, *arbeitet weitgehend lokal* an der gerade fokussierten Idee, bringt lediglich *an sehr wenigen Stellen neue Verknüpfungen* mit anderen Teilen des Problemlöseprozesses oder neue Ideen von außerhalb des Diskurses ein.

Typische Beispiele:

Hatten wir da nicht vorhin was dazu? War x nicht gerade? Ich glaube, dass hat immer was mit der Zahl zu tun, mit der man anfängt…

Nicht einschlägig: „Steht da nicht was im Skript dazu?“, „Ich kann mal die Definition nachsehen.“

Bemerkungen, Anmerkungen während der Kodiererschulung:

Allgemein: Nicht einschlägig ist lediglich ein Verweis auf andere Informationsquellen.

Qualität der Exploration von Hypothesen

Frage: Inwiefern werden vom Lerner verschiedene Vermutungen exploriert? Inwiefern wird vor Beginn oder während der Evidenzgenerierung die Plausibilität einer Vermutung kritisch hinterfragt?

Kriterien: Die Kodierung fokussiert insbesondere auf die Exploration und Generierung von Vermutungen. Von Interesse ist dabei, inwiefern verschiedene Vermutungen exploriert werden und ob die Plausibilität der Hypothese vor Beginn der (formalen) Evidenzgenerierung kritisch überprüft wird.

1	Der Lerner exploriert *verschiedene Vermutungen*, er versucht möglichst viele verschiedene Vermutungen zu finden. *Im Wesentlichen* wird die *Plausibilität seiner Vermutung* jeweils vor Beginn oder während der Evidenzgenerierung *kritisch hinterfragt*, indem (informale) Argumente für *und* gegen die Hypothese gesucht werden.
2	Der Lerner exploriert verschiedene Vermutungen, er versucht verschiedene Vermutungen zu finden. Die *Plausibilität* seiner Vermutung wird jeweils *nur teilweise kritisch* hinterfragt, indem (informale) Argumente für *oder* gegen die Hypothese gesucht werden.
3	Der Lerner exploriert *eine Vermutung*. Die Plausibilität dieser Vermutung wird *kritisch* hinterfragt, indem (informale) Argumente für oder gegen die Hypothese gesucht werden.
4	Der Lerner exploriert im Laufe des Problemlöseprozesses verschiedene Vermutungen. Die *Plausibilität* der Vermutungen wird jeweils vor Beginn oder während der Evidenzgenerierung *nicht kritisch* hinterfragt. An Stellen, an denen kein weiterer Fortschritt erfolgt, der Problemlöseprozess zu stagnieren beginnt, Schwierigkeiten auftreten oder Bearbeitungszeit über bleibt, wird eine *neue Vermutung exploriert*, die wiederum nicht kritisch hinterfragt wird.
5	Der Lerner exploriert im Laufe des Problemlöseprozesses eine Vermutung. Die Plausibilität einer Vermutung wird vor Beginn oder während der Evidenzgenerierung nicht kritisch hinterfragt. *Weitere Vermutungen werden nicht exploriert*.

Typische Beispiele:

Beispiele für eine kritische Prüfung der Plausibilität: „Wir sollten vielleicht auch mal versuchen ein Gegenbeispiel zu finden!“, „Ich glaube nicht, dass das immer gilt, da…(ein Argument als Grund für die Zweifel muss angegeben werden)“, „Das gilt jetzt erstmal nur für die Beispiele, aber wenn es allgemein gelten soll, dann muss gezeigt werden, dass….“, „Lass uns lieber noch weitere Beispiele ansehen!“, „Das wird aber schwer zu beweisen sein, da…(+Begründung)“

Als kritische Aktivitäten werden ebenfalls gewertet: *das Betrachten von Extremfällen sowie eine kritische Analyse der Struktur der zu beweisenden Behauptung.*

Bemerkungen, Anmerkungen während der Kodiererschulung:

Anmerkung zu Code 1: Um Code 1 zu vergeben, muss der Lerner vor Beginn und während der Evidenzgenerierung Argumente suchen, die die Plausibilität seiner Hypothese stützen (z.B. Exploration von Beispielen) **und** er muss sich zusätzlich darüber hinaus Gedanken machen, weshalb seine Hypothese nicht gelten könnte *(z.B.: „Lässt sich auch ein Gegenbeispiel finden?"*, *„ich glaube nicht, dass man das so allgemein aufschreiben kann, da…")*

Allgemein: Das Skizzieren einer Beweisskizze kann als kritische Exploration aufgefasst werden, wenn der Lerner sich z.B. Gedanken darüber macht, an welchen Stellen des Beweises Probleme auftreten können oder, wenn er sich z.B. darüber Gedanken macht, wie die zu beweisende Behauptung am Ende aussehen sollte (ggf. im entsprechendem Repräsentationssystem).

Allgemein: Werden verschiedene Vermutungen exploriert, so wird mindestens der Code 2 vergeben, wenn eine der Vermutung kritisch exploriert wurde.

Allgemein: Bei dieser Variable werden Vermutungen, die offensichtlich von „Nachbarn" stammen, nicht berücksichtigt.

Allgemein: Werden nach der Formulierung der Vermutung weitere Beispiele betrachtet, so wird dies lediglich als kritische Aktivität kodiert, wenn deutlich wird, dass der Lerner die Generierung von weiteren Beispielen dazu verwendet, sich selbst noch stärker von der Vermutung zu überzeugen.

Allgemein: Wird eine Behauptung lediglich vom Lernpartner übernommen und selbst überhaupt nicht in Hinblick auf deren Plausibilität untersucht, so wird ebenfalls Code 5 vergeben.

Qualität der Argumentstruktur

Frage: Inwiefern bringt der Lerner mathematische Schlüsse in den Diskurs ein, die in Bezug auf ihre Struktur vollständig sind? Inwiefern gibt der Lerner Hinweise darauf, auf welcher Basis und mit welcher Sicherheit ein Schluss gezogen wird, wo es im Kontext des Diskurses notwendig und hilfreich ist?

Kriterien: Betrachtet werden Argumente, die wesentlich für den mathematischen Problemlöseprozess sind. Relevant ist, inwiefern über eine Behauptung (*Claim*) hinaus weitere Strukturelemente von Argumenten expliziert werden:

Claim: Ein Claim ist eine Behauptung, also eine Aussage deren (angenommene) Gültigkeit kommuniziert wird.

Solche Behauptungen können in Argumentationsprozessen unterschiedliche Rollen (Status) erfüllen, z.B. als (vorläufige, nicht abgesicherte) Vermutung oder als (z.B. deduktiv abgesicherte) Schlussfolgerung.

Hinweise zur Identifizierung

Claims sind Aussagen, die im Kontext des Studiums begründungsbedürftig sind. Dazu zählen wir Aussagen, deren Gültigkeit nicht anhand einer einfachen elementaren Rechenoperation (mit Regeln aus dem Schulunterricht) belegt werden kann oder deren Gültigkeit aus dem Schulunterricht nicht bereits gut bekannt ist. So wird z.B. die Aussage *„2x+4=2(x+1)"* nicht als Claim kodiert, eine Aussage wie *„die Summe zweier gerader Zahlen, ist immer gerade"* jedoch schon. Behauptungen sind solche Aussagen, die im Rahmen des Mathematikstudiums eine zusätzliche Begründung/ Erklärung erfordern und deren Gültigkeit nicht einfach akzeptiert wird. Bem.: die Gültigkeit und Akzeptanz einer Aussage ist immer abhängig von der Community.

Der Status einer Behauptung gibt innerhalb eines diskursiven Prozesses Aufschluss darüber, ob es sich um eine Hypothese oder um eine abgesicherte Schlussfolgerung handelt, deren Gültigkeit bereits gesichert wurde (angelehnt an Duval, (2002)).

Datum: Gründe für die Behauptung aus dem Kontext des Problemlöseprozesses, also bereits etablierte Aussagen oder im Diskurs als korrekt angenommene Aussagen. Dies umfasst nicht Aussagen, die aus der Rahmentheorie (z.B. Vorlesung, Vorwissen,...) herangezogen werden.

Stützung: Gründe für die Behauptung (bzw. dafür, dass die Behauptung aus dem Datum folgt). Wir unterscheiden dabei Stützungen, die aus einer Rahmentheorie (z.B. Vorlesung, Vorwissen,...) herangezogen werden (d.h. formale Stützungen) sowie anschauliche oder empirische Stützungen, die auf Beispiele basieren (d.h. informale Stützungen).

Einschränkungen: Aussagen über die Sicherheit eines Schlusses (modal qualifier (Toulmin, 1996), epistemic status (Duval, 2002): wahrscheinlich, wenn ich mich nicht irre, sicher,...) bzw. Aussagen über mögliche einschränkende Bedingungen für die Gültigkeit des Schlusses („zumindest wenn x gerade ist").

Relevant ist jeweils nicht, ob die Aussagen, Stützungen und Schlüsse korrekt sind, sondern ob sie strukturell vollständig sind.

Zentral ist weiter nicht, dass jedes Argument vollständig ist, sondern dass die im Diskurs an der jeweiligen Stelle wesentlichen, aber nicht ohnehin anderweitig naheliegenden Teile expliziert werden.

1	Die Argumente des Lerners sind *durchgehend in Bezug auf ihre Struktur ausführlich* und geben, wo es im Kontext des Diskurses möglich und hilfreich ist, *Hinweise darauf, auf welcher Basis und mit welcher Sicherheit* welcher Schluss gezogen wird. Stützungen gehen *meist über empirische oder anschauliche Stützungen* hinaus.
2	Die Argumente des Lerners gehen *überwiegend über die Formulierung einer Behauptung* hinaus. Nur *an wenigen oder an wenig zentralen Stellen fehlen notwendige Hinweise darauf, auf welcher Basis oder mit welcher Sicherheit der Schluss gezogen wird.* Neben *empirischen oder anschaulichen Stützungen* finden sich *auch Stützungen,* die *aus der Rahmentheorie* herangezogen werden.
3	Die Argumente des Lerners gehen überwiegend über die Formulierung einer Behauptung hinaus. Dennoch *fehlen an wesentlichen oder vielen Stellen notwendige Hinweise darauf, auf welcher Basis oder mit welcher Sicherheit der Schluss gezogen wird.* Stützungen beschränken sich *zumeist auf empirische oder anschauliche Stützungen.*
4	Die Argumente des Lerners gehen *vereinzelt, aber nicht überwiegend über die Formulierung einer Behauptung hinaus*. An *vielen wesentlichen Stellen* fehlen notwendige

	Hinweise darauf, auf welcher Basis oder mit welcher Sicherheit der Schluss gezogen wird. Stützungen beschränken sich zumeist auf empirische oder anschauliche Stützungen.
5	Auch *bei zentralen Argumenten und solchen, wo es sich anbieten würde, beschränken sich Argumente des Lerners im Wesentlichen auf die Angabe einer Behauptung*, ggf. mit wenig aussagekräftigen Einschränkungen.

Typische Beispiele:

Beispiele dazu, an welchen Stellen (bzw. an welch analogen Stellen) wir unbedingt eine Stützung und ggf. einen Qualifier/ Einschränkung erwarten:

Behauptung: Das Ergebnis ist eine Quadratzahl.
*-> Eine Stützung ist notwendig: wenn empirisch (z.B.: 1*2*3*4*+1=5^2=25 und 2*3*4*5+1=11^2=121), dann muss ein Qualifier/ Einschränkung erfolgen (z.B.: Es scheint so, als sei das Ergebnis immer eine Quadratzahl. Das müssen wir jetzt aber noch beweisen). Die Einschränkung, dass die Behauptung noch zu beweisen ist, muss allerdings nicht unmittelbar folgen. Es reicht, wenn gefolgert wird, dass dies erstmal eine Vermutung ist.*

Behauptung: Das Ergebnis lässt sich schreiben als $(n(n+3)+1)^2$.
*-> Eine Stützung ist notwendig: wenn empirisch (z.B.: 1*4=5 ->5^2=25 und $(2*5+1)^2=11^2$=121), dann muss ein Qualifier/ Einschränkung erfolgen (z.B.: Dies gilt zumindest schon mal für die Beispiele, aber das ist noch nicht bewiesen).*

n(n+1)(n+2)(n+3)+1 = 2l(2l+1)(2l+2)(2l+3)+1= 2k+1mit k := l(2l+1)(2l+2)(2l+3), $k \epsilon \mathbb{Z}$

Behauptung: Das Ergebnis ist das Quadrat einer Primzahl.
-> Hier kann nur eine empirische Stützung erfolgen. Es muss explizit angemerkt werden, dass das erstmal nur für die Beispiele gilt und noch zu beweisen ist.

Bemerkungen, Anmerkungen während der Kodiererschulung:

Anmerkung: Es ist zu berücksichtigen, dass insbesondere Behauptungen, die lediglich auf Beispiele stützen, Hinweise, mit welcher Sicherheit der Schluss gezogen wurde, enthalten sollen.

Reaktion auf fachlich falsche Äußerungen des Lernpartners

Frage: Inwiefern trägt der Lerner zu einem kritisch-konstruktiven fachlichen Diskurs bei, indem er auf fachlich falsche Äußerungen des Lernpartners reagiert? Inwiefern erfolgen über die Kritik hinaus Begründungen und Alternativvorschläge?

Kriterien: Die Kodierung fokussiert auf fachlich falsche Äußerungen des Lernpartners, insbesondere auf Stellen, an denen eine Reaktion des Lerners stattfindet. Kodiert wird, ob eine Reaktion des Lerners auf eine fachlich falsche Äußerung des Lernpartners erfolgt und inwiefern über die Kritik hinaus Begründungen und Alternativvorschläge angeführt werden. Dabei ist weniger von Bedeutung, ob wirklich alle fachlichen Fehler des Lernpartners identifiziert werden (z.B. kleine formale Mängel, die die Bedeutung einer Aussage nicht wesentlich einschränken); wichtiger ist vielmehr, ob zentrale Fehler, die zu längeren Fehlwegen im Problemlöseprozess führen (könnten), erkannt werden und inwiefern eine elaborierte Kritik stattfindet.

1	*Fachlich falsche Äußerungen des Lernpartners* werden *häufig* vom Lerner *kritisch hinterfragt*. Die Kritik wird *durchgehend begründet und Alternativvorschläge* werden angeführt.
2	Fachlich falsche Äußerungen des Lernpartners werden häufig vom Lerner kritisch hinterfragt. Die Kritik wird *überwiegend begründet und nur an wenigen zentralen Stellen fehlen Alternativvorschläge*.
3	Fachlich falsche Äußerungen des Lernpartners werden *mehrfach* vom Lerner *kritisch hinterfragt*. Die Kritik wird *nur teilweise begründet* und *nur an wenigen zentralen Stellen werden Alternativvorschläge angeführt*, aber *mindestens einmal*.
4	Fachlich falschen Äußerungen des Lernpartners werden *nur vereinzelt* vom Lerner *kritisch hinterfragt*. Die Kritik wird ebenfalls *nur vereinzelt begründet, aber mindestens einmal*. *Alternativvorschläge werden nicht oder nur vereinzelt angeführt.*
5a	Fachlich falsche Äußerungen werden *überhaupt nicht oder nur vereinzelt erkannt*. *Begründungen für die Kritik oder Alternativvorschläge werden überhaupt nicht angeführt*.
5b	Der Lernpartner äußert bis auf einzelne Ausnahmen keine fachlich falschen Äußerungen. Kritik wäre somit auch nicht angebracht.

Typische Beispiele:

„Ich komm auf ein anderes Ergebnis, das müsste doch 11x^2 heißen. Ich glaube, du hast dich verrechnet“

„Das ist nicht 4x+5, schließlich sollen wir multiplizieren und nicht addieren“

Bemerkungen, Anmerkungen während der Kodiererschulung:

Anmerkung – zu einem bestimmten Fall: Der Lerner ist sich selbst an einer Stelle unsicher und fragt deshalb seinen Partner. Dieser antwortet ihm. Allerdings ist seine Antwort nicht korrekt. -> Auch wenn der Lerner selbst unsicher ist, wird dennoch von ihm erwartet, dass er die fehlerbehaftete Antwort seines Partner zumindest kritisch hinterfragt. Ist dies nicht der Fall, so wird dies negativ bewertet.

Qualität von fachlichen Fragen im Diskurs

Frage: Inwiefern prägt der Lerner den Diskurs durch sein Interesse, eine inhaltliche Abstimmung mit dem Lernpartner zu erreichen, indem er fachliche Fragen stellt? Inwiefern trägt der Lerner zu einer kollaborativen Zusammenarbeit bei, indem er die Perspektive des Lernpartners erfragt?

Kriterien: Berücksichtigt werden Äußerungen des Lerners an Stellen, an denen Unklarheiten, Zweifel oder unterschiedliche Sichtweisen auftreten, oder wo von den Lernern unterschiedliche Aspekte der Aufgabe bearbeitet werden. Hier wird insbesondere darauf fokussiert, inwiefern der Lerner die Möglichkeit nutzt, die Perspektive seines Lernpartners zu erfragen, um eine inhaltliche Abstimmung der Arbeit zu erreichen.

1	An vielen, geeigneten Stellen *(>3) trägt der Lerner weitgehend dazu bei*, eine *inhaltliche Abstimmung mit dem Lernpartner zu erreichen*, indem er (über rhetorische Fragen und Rückversicherungen hinaus) die *Perspektive sowie das Vorgehen des Lernpartners in regelmäßigen Abständen erfragt und einbezieht*. Die Fragen beziehen sich vorwiegend auf den gemeinsamen Arbeitsprozess oder auf die Beiträge des Lernpartners und tragen zur Weiterentwicklung gemeinsamer Ideen bei.
2	An einzelnen, geeigneten Stellen *(max. 3) versucht der Lerner (über rhetorische Fragen und Rückversicherungen hinaus)* eine inhaltliche Abstimmung mit dem Lernpartner zu erreichen, indem er *die Perspektive sowie das Vorgehen des Lernpartners erfragt und einbezieht*. Die Fragen beziehen sich vorwiegend auf den gemeinsamen Arbeitsprozess oder auf die Beiträge des Lernpartners.
3	Die inhaltliche Abstimmung zwischen den Lernpartnern beschränkt sich *im Wesentlichen auf mehrere (>3) rhetorische Fragen und Rückversicherungen für das eigene Vorgehen*. Die Perspektive sowie das Vorgehen des Lernpartners bleiben dabei weitgehend unberücksichtigt und werden nicht weiter einbezogen.
4	Die inhaltliche Abstimmung zwischen den Lernpartnern beschränkt sich *auf einzelne (max. 3) rhetorische Fragen und Rückversicherungen für das eigene Vorgehen*. Das Vorgehen des Lernpartners bleibt dabei völlig unberücksichtigt und wird überhaupt nicht hinterfragt.
5	*Versuche einer Abstimmung*, indem an wesentlichen Stellen an den Lernpartner Fragen gestellt werden, sind *nicht* zu erkennen. *Selbst rhetorische Fragen und Rückversicherungen* für das eigene Vorgehen treten *überhaupt nicht* auf.

Typische Beispiele:

Fragen, die über rhetorische Fragen sowie Rückversicherungen hinausgehen:

Bsp.: Sollen wir das nun formal aufschrieben? Ich versteh nicht, wie bist du darauf gekommen? Was wollten wir jetzt nochmal schreiben? Was ist deine Vermutung denn?

Rhetorische Fragen sowie Rückversicherungen:

Bsp.: (Lerner rechnet und sagt leise vor sich hin) Das müsste jetzt stimmen, oder? Passt? Du siehst auch noch nichts, oder?

Fragen bezüglich der Lösungsschritte im Programm werden nicht als fachliche Fragen gezählt, ebenso nicht wie rein organisatorische Fragen wie z.B. Schreibst du das auf oder soll ich?

Bemerkungen, Anmerkungen während der Kodiererschulung:

Wenn nur einmal eine Frage bezogen auf die Vorgehensweise des Lernpartners erfolgt und das Vorgehen nicht direkt einbezogen wird (also nur „Hast du schon eine Vermutung?“ und dann unmittelbar die eigene Vermutung vorgestellt wird), so wird dies mit max. 3 kodiert. Code 2 wird erst dann gegeben, wenn der Lerner auf die Antwort seines Partners direkt eingeht bzw. darauf anschließend Bezug nimmt.

Um Code 1 zu vergeben, ist es wichtig, dass Fragen, die sich auf den gemeinsamen Arbeitsprozess oder auf die Beiträge des Lernpartners beziehen, an verschiedenen Stellen auftreten.

Zyklen im Gespräch/ Inhaltliche Kohärenz

Frage: Inwiefern trägt der Lerner aktiv zu längeren Phasen eines zusammenhängenden, fachlichen Diskurses über die Aufgabenbearbeitung bei? Inwiefern trägt der Lerner durch geeignete Impulse dazu bei, gemeinsame Ideen weiterzuentwickeln.

Kriterien: Hier wird darauf fokussiert, inwiefern der Lerner eine gemeinsame Arbeit an einem geteilten Thema oder einer gemeinsamen Idee aktiv sicherstellt, sich daran beteiligt oder aber eine solche gemeinsame Arbeit vereinzelt oder wiederholt unterbricht.

Ein längerer gemeinsamer Diskurs zieht sich über wenigstens fünf bis sechs zusammenhängende Wortwechsel und ist von einem gemeinsamen Interesse an einem Thema geprägt.

1	Der Lerner trägt *wesentlich aktiv* dazu bei, dass *längere Phasen gemeinsamen Diskurses an einem gemeinsamen Thema* auftreten, beispielsweise indem er das Gespräch an geeigneten Stellen *durch geeignete Impulse fokussiert* und gemeinsame Ideen weiterentwickelt.
2	Der Lerner *beteiligt* sich durchgehend wesentlich an längeren Phasen gemeinsamen Diskurses an einem gemeinsamen Thema, beispielsweise indem er das Thema sowie *gemeinsame Ideen weiterentwickelt* und *nicht abschweift*.
3	Der Lerner *unterbricht vereinzelt* Phasen des gemeinsamen Diskurses, indem er vom gemeinsam diskutierten Thema oder der gemeinsamen Idee abschweift. *Überwiegend beteiligt er sich* jedoch auch an längeren Phasen des *gemeinsamen Diskurses*.

4	Der Lerner *unterbricht häufig* Phasen des gemeinsamen Diskurses, indem er vom gemeinsam diskutierten Thema oder der gemeinsamen Idee abschweift oder keinen Bezug auf Ideen des Lernpartners nimmt. Er *arbeitet weitgehend an seinen eigenen Ideen.*
5	Der Lerner *nimmt nur selten substantiellen Bezug auf die Äußerungen des Lernpartners*, sodass im Gespräch auch kürzere Phasen gemeinsamen Diskurses zu einem gemeinsamen Thema oder gemeinsame Ideen kaum auftreten.

Typische Beispiele:

Bsp.: „Lass uns das doch nochmal überprüfen!“; „Ich glaube wir können unsere Vermutung auch so schreiben…“; „ich habe es jetzt so aufgeschrieben, wie du vorgeschlagen hast, würde aber noch ergänzen, dass…“

Bemerkungen, Anmerkungen während der Kodiererschulung:

Wenn der Lerner einmal eine längere Phase abweicht, dann bereits 3

10.2 Ratingskalen

Videodatei:

Rater:

Durchgang:

Lerner – Code:

Ratings:

Var-Inhaltliche Korrektheit (Var_Inhalt):

Inwiefern bringt der Lerner fachlich korrekte Äußerungen in den Diskurs ein? Inwiefern werden also fachlich korrekte Aussagen über die betrachteten mathematischen Konzepte gemacht?

9	**5**	**4**	**3**	**2**	**1**

Var-Kognitives Niveau (Var_Kogn):

Inwiefern bringt der Lerner zu unterschiedlichen Phase des Argumentationsprozesses neue inhaltliche Ideen ein, stellt logische Beziehungen zwischen unterschiedlichen Ideen und Konzepten her und bringt somit neue Aspekte in den Diskurs ein?

9	**5**	**4**	**3**	**2**	**1**

Var-Exploration von Vermutungen (Var_Expl):

Inwiefern werden vom Lerner verschiedene Vermutungen exploriert? Inwiefern wird vor Beginn oder während der Evidenzgenerierung die Plausibilität einer Vermutung kritisch hinterfragt?

9	**5**	**4**	**3**	**2**	**1**

Var-Argumentstruktur (Var_Arg):

Inwiefern bringt der Lerner mathematische Schlüsse in den Diskurs ein, die in Bezug auf ihre Struktur vollständig sind? Inwiefern gibt der Lerner Hinweise darauf, auf welcher Basis und mit welcher Sicherheit ein Schluss gezogen wird, wo es im Kontext des Diskurses notwendig und hilfreich ist?

9	**5**	**4**	**3**	**2**	**1**

Var-Reaktion auf fachlich falsche Äußerungen (Var_Reak):

Inwiefern trägt der Lerner zu einem kritisch-konstruktiven fachlichen Diskurs bei, indem er auf fachlich falsche Äußerungen des Lernpartners reagiert? Inwiefern erfolgen über die Kritik hinaus Begründungen und Alternativvorschläge?

9	**5b**	**5a**	**4**	**3**	**2**	**1**

Var-Fachliche Fragen (Var_Frage):

Inwiefern prägt der Lerner den Diskurs durch sein Interesse, eine inhaltliche Abstimmung mit dem Lernpartner zu erreichen, indem er fachliche Fragen stellt? Inwiefern trägt der Lerner zu einer kollaborativen Zusammenarbeit bei, indem er die Perspektive des Lernpartners erfragt?

9	**5**	**4**	**3**	**2**	**1**

Var-Inhaltliche Kohärenz (Var_Zyklus):

Inwiefern trägt der Lerner aktiv zu längeren Phasen eines zusammenhängenden, fachlichen Diskurses über die Aufgabenbearbeitung bei? Inwiefern trägt der Lerner durch geeignete Impulse dazu bei, gemeinsame Ideen weiterzuentwickeln?

9	**5**	**4**	**3**	**2**	**1**

Notizen:

10.3 Rater-Training

Trainingsprinzip: wiederholte Herstellung einer Verknüpfung von beobachtbaren mathematischen Argumentationsprozessen mit der konzeptgemäßen Qualitätsstufe durch den übenden Rater. Fehler werden durch unmittelbares Feedback korrigiert.

Zeit	**Ziel der Phase**	**Methode**	**Hinweise**
Ca. 2,5 h	Die Rater sollen mit den inhaltlichen und formalen Anforderungen der Aufgabe vertraut werden und Prozessmerkmale mathematischen Argumentierens und Beweisens in kooperativen Settings kennen lernen.	Die Aufgabe und das Kodiermanual werden vorgelegt. Der Rater löst die Aufgabe zunächst selbst und liest das Kodiermanual durch. Der Rater wird anschließend mündlich über die Herkunft der Konzepte und Prozessmerkmale mathematischen Argumentierens und Beweisens aufgeklärt. Die Prozessmerkmalsdefinitionen und Beschreibungen sowie die Verankerungen der Qualitätsstufen durch Beispiele werden gemeinsam diskutiert. Fragen können dabei jederzeit gestellt werden.	
Ca. 3,5 h	Ziel ist das Erkennen von (Skalenausprägungs-) Unterschieden: Die Rater sollen ein Gefühl für die Qualität der mathematischen Argumentationskompetenz von StudienanfängerInnen in kooperativen Settings entwickeln.	*Diskriminationstraining*: 3 Videos (ggf. 3 Videoausschnitte) werden vorgelegt und sind vom Rater in eine Rangfolge hinsichtlich ihrer Qualitätsausprägung zu bringen (für jedes Prozessmerkmal einzeln). Anschließend werden die Ergebnisse im Plenum diskutiert und begründet.	Die Videos sollten sich im Hinblick auf die Qualität der Prozessmerkmale deutlich voneinander unterscheiden. Möglichst eine sehr erfolgreiche und eine weniger erfolgreiche Dyade. Es ist darauf zu achten, dass möglichst viele verschiedene

			Qualitätsausprägungen in vielfältiger Kombination auftreten.
Ca. 30 min	Die Rater sollen ein Gefühl für positive und negative mathematische Argumentationsprozesse in kooperativen Settings entwickeln.	*Herstellungstraining*: Ein Video mittlerer Qualität (aus dem Diskriminationstraining bekannt) wird vorgelegt. Der Rater soll jetzt dieses Video in eine positive und negative Version umwandeln und seine Gedanken dabei erläutern.	Das Video wurde bereits im Rahmen des Diskriminationstrainings angesehen.
Ca. 1,5 h (nur bei zusätzlichen Bedarf)	Die Rater sollen lernen, mathematische Argumentationskompetenz anhand verschiedener Prozessmerkmale zu beurteilen. Überlappungen und Verwechslungen zwischen diesen Merkmalen sind zu vermeiden.	*Konzept-Unterscheidungstraining*: Ein Video wird vorgelegt und der Rater ordnet aufgrund seines (gelernten) Verständnisses jedem Prozessmerkmal ein typisches Beispiel (typische Argumentationsprozesse) aus dem Video zu. Die Zuordnung muss jeweils begründet werden.	
Ca. 1,5 h	Die Rater sollen ein gemeinsames Verständnis der einzelnen Prozessmerkmale mathematischen Argumentierens entwickeln und an Erfahrung gewinnen.	Ein weiteres Video wird vorgelegt und ist vom Rater hinsichtlich aller Prozessmerkmale zu raten. Während des Ratings werden Fragen und Schwierigkeiten notiert, die anschließend im Plenum diskutiert werden. Bei Bedarf wird das Kodiermanual adaptiert und es werden zusätzliche Regeln eingeführt.	

Anschließend werden weitere Videos geratet und statistisch analysiert (z.B. Häufigkeiten, Prozentuale Übereinstimmung, Cohens Kappa). Die Ergebnisse bilden die Grundlage für die Trainingsphase 2.

Zeit	**Ziel der Phase**	**Methode**	**Hinweise**
Ca. 30 min	Jeder Rater soll seine Einschätzung mit den Einschätzungen der anderen Rater vergleichen. Die Interrater-Übereinstimmung wird überprüft.	Die Interrater-Übereinstimmung für alle Prozessmerkmale wird diskutiert, insbesondere Merkmale mit einer geringen Übereinstimmung.	
Ca. 15 min	Die Einschätzung für verschiedene Qualitätsausprägungen soll trainieret werden.	Qualitätsausprägungen, die sehr häufig oder nur sehr selten vergeben wurden, werden diskutiert.	
Ca. 40 min	Ursachen für Unklarheiten und für eine geringe Interrater-Reliabilität werden analysiert und diskutiert.	Extremfälle (besonders auffällige Argumentationsprozesse sowie Fälle mit besonders geringer Übereinstimmung) werden gemeinsam analysiert. Die entsprechenden Videosequenzen werden erneut gemeinsam angesehen.	

Das Training kann dann beendet werden, wenn alle Rater ein gemeinsames theoretisches Verständnis zu den Prozessmerkmalen mathematischen Argumentierens und Beweisens aufgebaut haben (zeigt sich anhand einer guten Interrater-Reliabilität) und alle Fragen geklärt werden konnten. Alle Videos, die für das Training verwendet wurden, sollten nicht in die Datenanalyse eingebracht werden.

(Langer & Schulz von Thun, 2007; Wiesbeck, 2015)

10.4 Kodier-Manual der Eigenlösungen als Produkt kooperativer Argumentationsprozesse

Inhalt der Vermutung Hier geht es darum, welche Vermutung gefunden wurde (formale Mängel spielen hierbei noch keine Rolle), also um den Inhalt der Vermutung. Solange weitgehend zweifelsfrei erkennbar ist, was gemeint ist (auch bei formalen Mängeln im Aufschreiben) wird die Vermutung interpretiert.		
Code	Beschreibung	Beispiel
<u>*0x Vermutung falsch oder nicht interpretierbar*</u> Es wurde keine korrekte Vermutung gefunden oder formuliert oder die Vermutung ist nicht interpretierbar.		
0 Keine Vermutung	Es wurde keine Vermutung explizit formuliert. Dieser Code wird auch dann vergeben, wenn zwar aus dem Beweis die Vermutung ersichtlich wird, diese aber nicht explizit aufgeschrieben wurde.	
1 Falsche Vermutung	Es wurde eine nicht korrekte Vermutung formuliert.	„Vermutung: Das Ergebnis ist immer das Quadrat einer Primzahl.“
2 Nicht-interpretierbare Vermutung	Es wurde eine auch bei guten Willen nicht interpretierbare Vermutung formuliert. Die Vermutung ist aus inhaltlicher Sicht nicht verständlich.	„ z.z.: $x^2 \wedge x\|1 \wedge x$ ungerade“,…
<u>*1x Vermutung korrekt, aber teilweise trivial*</u> Es wurde mindestens eine korrekte, aber triviale Vermutung gefunden und formuliert, d.h. es könnte eine stärke Vermutung abgeleitet werden.		
0 Korrekte, teilweise triviale Vermutung	Es wurde eine korrekte Vermutung formuliert. Allerding ist diese eher von trivialer Natur, d.h. eine stärkere Vermutung hätte abgeleitet werden können.	„Das Ergebnis ist immer ungerade“, „ohne 1 ergibt sich eine gerade Zahl“, „x(x+1)(x+2)(x+3) ist immer gerade.“…

1 Mehrere korrekte, teilweise trivialen Vermutungen	Es wurden mehrere korrekte Vermutungen formuliert. Allerding sind diese alle eher von trivialer Natur, d.h. eine stärkere Vermutung hätte abgeleitet werden können.	„Wenn man vier aufeinanderfolgende Zahlen miteinander multipliziert und 1 addiert, dann ergibt sich immer eine ungerade Zahl und ohne 1 hat man eine gerade Zahl“
2 Korrekte, teilwiese triviale Vermutung, implizit	Es wurde eine korrekte Vermutung nur implizit formuliert. Diese implizite Vermutung ist eher von trivialer Natur, d.h. eine stärkere Vermutung hätte abgeleitet werden können.	=> ungerade
3 Korrekte, nicht-triviale Vermutung & falsche Vermutung	Es wurde eine nicht-triviale, korrekte Vermutung und eine falsche Vermutung formuliert. Der Fokus liegt auf der nicht-trivialen, korrekten Vermutung.	
<u>*2x Vermutung korrekt und nicht trivial*</u> Es wurde mindestens eine korrekte und nicht triviale Vermutung gefunden und formuliert.		
0 Korrekte und nicht triviale Vermutung	Es wurde eine korrekte, nicht triviale Vermutung formuliert.	„Wenn man vier aufeinanderfolgende ganze Zahlen miteinander multipliziert und 1 addiert, dann ergibt sich immer eine Quadratzahl.“,… „Wenn man vier aufeinanderfolgende ganze Zahlen multipliziert und 1 addiert, dann bleibt bei Division durch 4 Rest 1.

1 Mehrere korrekte Vermutungen und davon eine nicht triviale Vermutung	Es wurden mehrere korrekte Vermutungen formuliert. Eine davon ist nicht trivial.	„Wenn man vier aufeinanderfolgende Zahlen miteinander multipliziert und 1 addiert, dann ergibt sich immer eine ungerade Zahl. Diese Zahl ist immer eine Quadratzahl.“
2 Korrekte und nicht triviale Vermutung, implizit	Es wurde eine korrekte, nicht triviale Vermutung nur implizit formuliert.	=> Das Ergebnis ist eine Quadratzahl
Form der Vermutung Hier geht es darum, ob die Vermutung formal und verbal klar und eindeutig formuliert ist – unabhängig davon ob sie inhaltlich korrekt und/oder trivial ist oder nicht. Dabei geht es um die Verwendung von Variablen, Junktoren, Quantoren, Implikationspfeilen und Fachsprache.		
Code	Beschreibung	Beispiel
0x Vermutung mit gravierenden formalen Mängeln Es wurde keine formal verständliche Vermutung (unabhängig vom Inhalt der Vermutung) gefunden oder formuliert.		
0 Keine Vermutung oder nur implizite Vermutung	Es wurde keine Vermutung explizit formuliert. Dieser Code wird auch dann vergeben, wenn zwar aus dem Beweis die Vermutung ersichtlich wird, diese aber nicht explizit aufgeschrieben wurde.	
1 Vermutung mathematisch nicht korrekt formuliert, gravierende Mängel, überhaupt nicht interpretierbar	Es wurde eine Vermutung formuliert. Die Formulierung (unabhängig vom Inhalt der Vermutung!) ist jedoch nicht korrekt. Die Mängel sind so gravierend, dass die Aussage keinen Sinn macht und somit inhaltlich nicht interpretierbar ist.	z.B.: dieselbe Variable wird für unterschiedliche algebraische Ausdrücke verwendet oder mehrere formale Mängel treten gleichzeitig auf; „z.z.: $x^2 \wedge x|1 \wedge x$ ungerade“
2	Es wurde eine Vermutung formuliert. Die Formulierung	z.B.: dieselbe Variable wird für unterschiedliche

Vermutung mathematisch nicht korrekt formuliert, gravierende Mängel, nur schwer verständlich	(unabhängig vom Inhalt der Vermutung!) ist jedoch nicht korrekt. Die Vermutung ist inhaltlich interpretierbar, aber aufgrund der gravierenden formalen Mängel nur sehr schwer verständlich.	algebraische Ausdrücke verwendet oder mehrere formale Mängel treten gleichzeitig auf; „z.z. ∃x∈Z: x^2|x(x+1)(x+2)(x+3)+1“,…
<u>1x Vermutung mit kleinen formalen Mängeln</u> Es wurde eine formal verständliche, allerdings nicht ganz formal korrekte Vermutung (unabhängig vom Inhalt der Vermutung) gefunden und formuliert. Kleine oder wenige formale Mängel sind zu finden, die jedoch die Bedeutung der Vermutung kaum einschränken.		
0 Vermutung mathematisch nicht ganz korrekt formuliert, nur wenige und kleine formale Mängel, die die Bedeutung der Aussage kaum einschränken	Es wurde eine Vermutung formuliert, die formale Mängel enthält. Die Vermutung ist aber im Prinzip verständlich. Allerdings finden sich kleine formale Mängel, so dass die Vermutung nicht vollständig sauber aufgeschrieben wurde. Die formalen Mängel sind so geringfügig, dass sie die Bedeutung der Aussage kaum in ihrer Verständlichkeit einschränken.	Variablen werden nicht systematisch eingeführt (keine Angaben zum Definitionsbereich), statt Verwendung des Allquantors wird versehentlich der Existenzquantor verwendet,…
1 Vermutung nicht ganz korrekt formuliert, Verknüpfung zwischen Voraussetzung und Behauptung fehlt	Es wurde eine Vermutung formuliert, die formale Mängel enthält. Der Wesentliche formale Mangel besteht darin, dass keine Verknüpfung zwischen Voraussetzung und Behauptung hergestellt wird. Darüber hinaus finden sich weitgehend keine weiteren formalen Fehler.	„Das Ergebnis ist immer eine Quadratzahl.“
<u>2x Vermutung formal korrekt</u> Es wurde eine formal verständliche und formal korrekte Vermutung (unabhängig vom Inhalt der Vermutung) gefunden und formuliert.		

0 Vermutung formal korrekt und verständlich formuliert	Es wurde eine verständliche und weitgehend formal korrekte Vermutung aufgeschrieben. Kleinere Probleme wie offensichtliche Schreibfehler werden toleriert.	z.z.: „∀x∊Z: x(x+1)(x+2)(x+3)+1 = (x²+(3x+1))², „∀x∊Z mit x(x+1)(x+2)(x+3)+1 ergibt sich eine Quadratzahl;…
Vorhandene Beweisideen Hier geht es darum, ob die wesentlichen Ideen des Beweises erkennbar sind (wesentlichen zentralen Ideen siehe Erklärung zur Aufgabe). Es geht dabei weniger darum, ob der Beweis formal korrekt dargelegt ist. Zentral ist, dass die wesentlichen Ideen inhaltlich erkennbar sind.		
Code	Beschreibung	Beispiel
0x Weniger als die Hälfte der zentralen Ideen Es wurden weniger als die Hälfte der zentralen Ideen angesprochen.		
0 Keine zentralen Ideen	Es wurden keine zentralen Ideen angesprochen. Entweder wurde nicht versucht, eine Vermutung zu beweisen oder es wurde keine zentrale Idee gefunden. Hierzu zählt auch ein reines Betrachten von Beispielen.	s. Erklärung zur Aufgabe
1 Weniger als die Hälfte der zentralen Ideen	Es wurden weniger als die Hälfte der zentralen Ideen angesprochen.	s. Erklärung zur Aufgabe
1x Mehr als die Hälfte der zentralen Ideen Es wurden mehr als die Hälfte der zentralen Ideen im korrekten Zusammenhang angesprochen.		
0 Mehr als die Hälfte der zentralen Ideen	Es wurden mehr als die Hälfte der zentralen Ideen angesprochen, aber nicht alle Ideen gefunden. Einige Ideen fehlen oder sind nicht nachvollziehbar.	s. Erklärung zur Aufgabe
2x Alle zentralen Ideen		

Es wurden alle zentralen Ideen angesprochen.		
0 Alle zentralen Ideen	Es wurden alle zentralen Ideen. Der Beweis enthält die wesentlichen Ideen.	s. Erklärung zur Aufgabe
Form des Beweises Hier geht es darum, ob die im Beweis erkennbaren Argumente (unabhängig von der inhaltlichen Korrektheit der Argumente) formal verständlich und formal korrekt dargelegt sind. Es geht hier lediglich um die *Darstellung der Argumente*, unabhängig von deren inhaltlicher Korrektheit. Beispielsweise wären formal perfekt dargelegte, beispielbasierte Argumente für Allaussagen durchaus mit „2x“ zu kodieren. Es geht im Wesentlichen darum, ob für die einzelnen Schlüsse im Beweis Variablen, Junktoren, Quantoren und Fachsprache adäquat verwendet wurde.		
Code	Beschreibung	Beispiel
0x Argumente mit gravierenden formalen Mängeln Es wurden keine formal korrekt dargelegten Argumente gefunden oder es wurden Argumente mit gravierenden formalen Mängeln formuliert, die wesentliche Teile des Beweises unakzeptabel machen.		
0 Keine Argumente	Es wurden keine Argumente formuliert.	
1 Argumente mit gravierenden formalen Mängeln	Es wurden Argumente mit gravierenden formalen Mängeln formuliert, die bei ernsthafter Interpretation alle wesentlichen Teile des Beweises unakzeptabel machen.	Die Art wie formale Notationen/ Variablen/ Quantoren/ Logische Junktoren verwendet werden ist bis auf wenige Ausnahmen kaum nachvollziehbar/häufig falsch/häufig mehrdeutig.
1x Argumente mit kleinen formalen Mängeln Es wurden im Wesentlichen interpretierbare, allerdings nicht ganz formal korrekte Argumente (unabhängig vom Inhalt der Argumente) gefunden und formuliert. Kleine oder wenige formale Mängel sind zu finden, die jedoch nur wenige Schritte des Beweises beeinflussen.		
0 Argumente mit wenigen oder leichten Mängeln	Es wurden Argumente mit wenigen oder leichten formalen Mängeln formuliert, die auch bei	Die Art wie formale Notationen/ Variablen/ Quantoren/ Logische

	ernsthafter Interpretation nur wenige Schritte des Beweises betreffen.	Junktoren verwendet werden, ist schwer nachvollziehbar/teilweise falsch/teilweise mehrdeutig. z.B. werden Variablen nicht korrekt eingeführt, sonst aber konsistent verwendet; vereinzelte falsche Verwendung des Gleichheitszeichens/ von Quantoren/ des Implikationspfeils.
2x Argumente weitgehend formal korrekt Es wurden formal verständliche und weitgehend formal korrekte Argumente (unabhängig vom Inhalt der Argumente) gefunden und formuliert. Keine oder nur minimale formale Mängel sind zu finden, die den Beweis nicht wesentlich beeinflussen.		
0 Argumente mit minimalen formalen Mängeln	Es wurden Argumente mit minimalen formalen Mängel formuliert, die die Interpretation des Beweises auch bei ernsthafter Interpretation nicht wesentlich beeinflussen	z.B.: Nicht für alle Variablen wird der Definitionsbereich angeben
1 Argumente formal korrekt	Es wurden formal korrekte Argumenten formuliert. Keine formalen Mängel sind zu finden.	Alle formalen Notationen werden in nachvollziehbarer Art verwendet.
Beweisstruktur/- kette Hier geht es darum, inwiefern klar aufeinander folgende Schritte aneinandergereiht (beginnend bei den Voraussetzungen, endend bei der Behauptung) werden und inwiefern Lücken zwischen den einzelnen Beweisschritten vorzufinden sind. Ausschlaggebend ist hier nicht die notierte Reihenfolge der Beweisschritte sondern die Organisation der präsentierten Argumente.		
Code	Beschreibung	Beispiel
0x Beweisstruktur/-kette mit gravierenden Mängeln Es wurden zahlreiche Teilschritte nicht bewiesen und/ oder es finden sich Fehler in der Beweisstruktur (d.h. die Struktur der Einzelschlüsse belegt nicht die Vermutung auf der Basis der		

Voraussetzungen und bekannter Sätze und Definitionen), wobei die gewählte Vorgehensweise nicht als heuristisches Hilfsmittel zu betrachten ist.		
0 Keine Beweisschritte	Es wurden keine Argumente formuliert.	
1 Fehler in der Beweiskette: Zahlreiche Teilschritten nicht beweisen	Es wurden zahlreiche Teilschritte nicht bewiesen, die in der Argumentation verwendet wurden, so dass der Beweis große Lücken aufweist.	„$x(x+1)(x+2)(x+3)+1= (x^2+x)\ (x^2+5x+6)= c^2$“
2 Fehler in der Beweisstruktur: die Struktur der Einzelschlüsse belegt nicht die Vermutung auf der Basis der Voraussetzungen und bekannter Sätze und Definitionen	Es wurden aufeinanderfolgende Schritte nicht klar aneinandergereiht. Die Argumentation ist gekennzeichnet durch eine nicht-korrekte Beweisstruktur.	Die Argumentation in der Bearbeitung geht von der Behauptung aus und leitet aus dieser unter Nutzung der Voraussetzung und bereits bekannter Definitionen, Sätze und Regeln eine weitere Aussage ab. Es wird versucht, etwas vom Typ „1=1“ abzuleiten. Der Beweis beginnt nicht bei der Voraussetzung und endet bei der Behauptung.
3 Fehler in der Beweiskette und der Beweisstruktur	Kombination aus Code 02 und 01.	„$x(x+1)(x+2)(x+3)+1= c^2$ $(x^2+x)\ (x^2+5x+6)= c^2$“,..“
1x Beweisstruktur/-kette mit kleinen Mängeln Es wurden nur wenige Teilschritte nicht bewiesen oder es wurde die Behauptung und Voraussetzung als heuristisches Hilfsmittel genutzt.		
0 Wenige Teilschritte nicht bewiesen	Es wurden die meisten Teilschritte, die in der Argumentation verwendet wurden, bewiesen. Nur vereinzelt finden sich noch Teilschritte, die nicht bewiesen	„$x(x+1)(x+2)(x+3)+1= (x^2+x)\ (x^2+5x+6) = (x^2+(3x+1))^2$

<table>
<tr><td></td><td>wurden, so dass der Beweis minimale Lücken aufweist.</td><td></td></tr>
<tr><td>1

Nutzung von Voraussetzung und Behauptung als potentielles heuristisches Mittel</td><td>Es wurde die Behauptung und Voraussetzung als heuristisches Hilfsmittel genutzt. Es wurde also anders als logisch konsistent vorgegangen, aber dieses Vorgehen wurde als heuristisches Mittel genutzt.</td><td>Die Argumentation in der Bearbeitung beginnt mit der Behauptung und leitet aus dieser unter Nutzung der Voraussetzung und bereits bekannter Definitionen, Sätze und Regeln eine weitere Aussage ab, die als strategische Hilfe dient und genutzt wird.</td></tr>
<tr><td colspan="3"><u>2x Beweisstruktur/-kette korrekt</u>

Es wurden alle aufeinanderfolgenden Schritte klar aneinandergereiht und bewiesen, so dass weitgehend keine Mängel in der Beweisstruktur/-kette vorliegen. Die Argumentation beginnt bei der Voraussetzung und endet bei der Behauptung bzw. bildet die Struktur, dass sie bei der Behauptung enden würde, wenn der Beweis zu Ende geführt worden wäre.</td></tr>
<tr><td>0</td><td>Es wurden alle aufeinanderfolgenden Schritte, die in der Argumentation verwendet wurden, klar aneinandergereiht und bewiesen, so dass weitgehend keine Mängel in der Beweisstruktur/- kette vorliegen.</td><td></td></tr>
</table>

List of figures

List of tables

11 References

Abel, C. F. (2003). Heuristics and problem solving. *New Directions for Teaching and Learning*, *2003*, 53–58.

Abrahamson, D., & Lindgren, R. (2014). Embodiment and embodied design. *The Cambridge Handbook of the Learning Sciences*, *2*, 358–376.

Alcock, L. (2010). Mathematicians' perspectives on the teaching and learning of proof. *Research in Collegiate Mathematics Education VII*, 63–92.

Alcock, L., & Inglis, M. (2008). Doctoral students' use of examples in evaluating and proving conjectures. *Educational Studies in Mathematics*, *69*, 111–129. https://doi.org/10.1007/s10649-008-9149-x

Alcock, L., & Weber, K. (2010). Undergraduates' example use in proof construction: Purposes and effectiveness. *Investigations in Mathematics Learning*, *3*, 1–22. https://doi.org/10.1080/24727466.2010.11790298

Alibert, D., & Thomas, M. (1991). *Advanced Mathematical Thinking*: Kluwer Dordrecht.

Alrø, H., & Skovsmose, O. (2002). Dialogue and Learning in Mathematics Education: Intention, Reflection. *Critique*, *29*.

Andriessen, J. E. B., & Schwarz, B. B. (2009). Argumentative design. In M.N. Muller & AN. Perret-Clermont (Eds.), *Argumentation and education* (pp. 145–174). Springer, Boston, MA.

Anjewierden, A., Gijlers, H., Kolloffel, B., Saab, N., & Hoog, R. de. (2011). Examining the relation between domain-related communication and collaborative inquiry learning. *Computers & Education*, *57*, 1741–1748. https://doi.org/10.1016/j.compedu.2011.03.010

Antonini, S. (2011). Generating examples: Focus on processes. *ZDM*, *43*, 205–217.

Artigue, M. (2001). What can we learn from educational research at the university level? In D. Holton, M. Artigue, U. Kirchgräber, J. Hillel, M. Niss, A. Schoenfeld (Eds.), *The teaching and learning of mathematics at university level* (pp. 207–220). New ICMI Study Series, vol 7. Springer, Dordrecht.

Asterhan, C. S. C., & Schwarz, B. B. (2007). The effects of monological and dialogical argumentation on concept learning in evolutionary theory. *Journal of Educational Psychology*, *99*, 626–639. https://doi.org/10.1037/0022-0663.99.3.626

Asterhan, C. S. C., & Schwarz, B. B. (2009). Argumentation and explanation in conceptual change: Indications from protocol analyses of peer-to-peer dialog. *Cognitive Science*, *33*, 374–400. https://doi.org/10.1111/j.1551-6709.2009.01017.x

Bagchi, A., & Wells, C. (1998). On the communication of mathematical reasoning. *Problems, Resources, and Issues in Mathematics Undergraduate Studies*, *8*, 15–27.

Balacheff, N. (1988). A study of students' proving processes at the junior high school level. *Second UCSMP International Conference on Mathematics Education*.

Balacheff, N. (1991). The benefits and limits of social interaction: The case of mathematical proof. In *Mathematical knowledge: Its growth through teaching* (pp. 173–192). Springer.

Balacheff, N. (1999). Is Argumentation an Obstacle? Invitation to a Debate. I*nternational Newsletter on the Teaching and Learning of Mathematical Proof*, 5-6. www.lettredelapreuve.itlResumeslBoerolBoero99.html (zuletzt aufgerufen am 10.11.2018).

Barron, B. (2000). Achieving coordination in collaborative problem-solving groups. *The Journal of the Learning Sciences*, *9*, 403–436. https://doi.org/10.1207/S15327809JLS0904_2

Barron, B. (2003). When smart groups fail. *The Journal of the Learning Sciences*, *12*, 307–359.

Blanton, M. L., & Stylianou, D. A. (2014). Understanding the role of transactive reasoning in classroom discourse as students learn to construct proofs. *The Journal of Mathematical Behavior*, *34*, 76–98. https://doi.org/10.1016/j.jmathb.2014.02.001

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

Boero, P. (1999). Argumentation and mathematical proof: A complex, productive, unavoidable relationship in mathematics and mathematics education. *International Newsletter on the Teaching and Learning of Mathematical Proof*, *7*. www.lettredelapreuve.org/OldPreuve/Newsletter/990708Theme/990708ThemeUK.html

Boero, P., Garuti, R., & Lemut, E. (2007). Approching theorems in grade VIII. *Theorems in School: from History, Epistemology and Cognition to Classroom Practice*, *249*, 249–264.

Cai, J., Mamona-Downs, J., & Weber, K. (2005). Mathematical problem solving: What we know and where we are going. *The Journal of Mathematical Behavior, 24(3-4), 217-220.*

http://dx.doi.org/10.1016/j.jmathb.2005.09.014

Carlson, M. P., & Bloom, I. (2005). The cyclic nature of problem solving: An emergent multidimensional problem-solving framework. *Educational Studies in Mathematics*, *58*, 45–75.

Carroll, C. D. (1977). The relative effectiveness of three geometric proof construction strategies. *Journal for Research in Mathematics Education*, 62–67.

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research*, *1*, 245–276.

Chazan, D. (1993). High school geometry students' justification for their views of empirical evidence and mathematical proof. *Educational Studies in Mathematics*, *24*, 359–387.

Chi, M. T. H. (2009). Active-constructive-interactive: A conceptual framework for differentiating learning activities. *Topics in Cognitive Science*, *1*, 73–105. https://doi.org/10.1111/j.1756-8765.2008.01005.x

Chi, M. T. H., & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist*, *49*, 219–243. https://doi.org/10.1080/00461520.2014.965823

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*, 865–887.

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, *6*, 284.

Clark, D. B., Sampson, V., Weinberger, A., & Erkens, G. (2007). Analytic frameworks for assessing dialogic argumentation in online learning environments. *Educational Psychology Review*, *19*, 343–374. https://doi.org/10.1007/s10648-007-9050-7

Clark, H. H., & Schaefer, E. F. (1989). Contributing to discourse. *Cognitive Science*, *13*, 259–294. https://doi.org/10.1016/0364-0213(89)90008-6

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*, 306–323. https://doi.org/10.1080/14794802.2014.950687

Clausen, M. (2002). *Unterrichtsqualität: Eine Frage der Perspektive? Empirische Analysen zur Übereinstimmung, Konstrukt- und Kriteriumsvalidität*. Münster: Waxmann Verlag.

Clausen, M., Reusser, K., & Klieme, E. (2003). Unterrichtsqualitat auf der Basis hoch-inferenter Unterrichtsbeurteilungen: Ein Vergleich zwischen Deutschland und der deutschsprachigen Schweiz. *Unterrichtswissenschaft*, *31*, 122–141.

Coe, R., & Ruthven, K. (1994). Proof practices and constructs of advanced mathematics students. *British Educational Research Journal*, *20*, 41–53.

Cohen, E. G. (1994). Restructuring the classroom: Conditions for productive small groups. *Review of Educational Research*, *64*, 1–35. https://doi.org/10.3102/00346543064001001

Common Core State Standards Initiative. (2010). National Governors Association Center for Best Practices and Council of Chief State School Officers. *Retrieved from http://www.corestandards.org.*

Corte, E. de, 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). Elsevier.

Dawkins, P. C. (2012). Extensions of the semantic/syntactic reasoning framework. *For the Learning of Mathematics*, *32*, 39–45.

Dawkins, P. C. (2015). Explication as a lens for the formalization of mathematical theory through guided reinvention. *The Journal of Mathematical Behavior*, *37*, 63–82.

Dawkins, P. C., & Karunakaran, S. S. (2016). Why research on proof-oriented mathematical behavior should attend to the role of particular mathematical content. *The Journal of Mathematical Behavior*, *44*, 65–75.

Dawkins, P. C., & Weber, K. (2017). Values and norms of proof for mathematicians and students. *Educational Studies in Mathematics*, *95*, 123–142.

Dieter, M. (2012). S*tudienabbruch und Studienfachwechsel in der Mathematik: Quantitative Bezifferung und empirische Untersuchung von Bedingungsfaktoren.* Dissertation, Universität Duisburg-Essen. http://duepublico.uni-duisburg-essen.de/servlets/DerivateServlet/Derivate-30759/Dieter_Miriam.pdf.

Douek, N. (1999). Argumentation and conceptualization in context: a case study on sunshadows in primary school. *Educational Studies in Mathematics*, *39*, 89–110.

Douek, N. (2007). Some remarks about argumentation and proof. *Theorems in School: from History, Epistemology and Cognition to Classroom Practice*, 163–181.

Dreyfus, T. (2002). Advanced mathematical thinking processes. In Tall D. (Ed.), *Advanced Mathematical Thinking.* Mathematics Education Library, vol 11. Springer, Dordrecht.

Dubinsky, E., & Yiparaki, O. (2000). On student understanding of AE and EA quantification. *Research in Collegiate Mathematics IV*, 239–289.

Ellis, A. B., Lockwood, E., Dogan, M. F., & Williams, C. C. W. (2013). Choosing and using examples: How example activity can support proof insight. In A. Lindmeier & A. Heinze (Eds.), *Proceedings of the 37th Conference of the International Group for the Psychology of Mathematics Education* (Vol. 2, pp. 265–272). Kiel, Germany.

Ellis, A. B., Ozgur, Z., Vinsonhaler, R., Dogan, M. F., Carolan, T., Lockwood, E., . . . Zaslavsky, O. (2017). Student thinking with examples: The criteria-affordances-purposes-strategies framework. *The Journal of Mathematical Behavior.* Advance online publication. https://doi.org/10.1016/j.jmathb.2017.06.003

Engelbrecht, J. (2010). Adding structure to the transition process to advanced mathematical activity. *International Journal of Mathematical Education in Science and Technology*, *41*, 143–154. https://doi.org/10.1080/00207390903391890

Epp, S. S. (2003). The role of logic in teaching proof. *The American Mathematical Monthly*, *110*, 886–899. https://doi.org/10.1080/00029890.2003.11920029

Epp, S. S. (2009). Proof issues with existential quantification. 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. 1, pp. 154-159).

Ernest, P. (1998). A postmodern perspective on research in mathematics education. In A. Sierpinska; J. Kilpatrick (Ends.), *Mathematics education as a research domain: A search for identity* (pp. 71–85). Dordrecht: Kluwer Academic Publishers.

Fischer, F., Kollar, I., Ufer, S., Sodian, B., Hussmann, H., Pekrun, R., . . . Fischer, M. (2014). Scientific Reasoning and Argumentation: Advancing an Interdisciplinary Research Agenda in Education. *Frontline Learning Research*, *2*, 28–45.

Fukawa-Connelly, T., Weber, K., & Mejía-Ramos, J. P. (2017). Informal Content and Student Note-Taking in Advanced Mathematics Classes. *Journal for Research in Mathematics Education*, *48*, 567–579.

Furinghetti, F., Maggiani, C., & Morselli, F. (2013). How mathematics students perceive the transition from secondary to tertiary level with particular reference to proof. *NORDISK MATEMATIKKDIDAKTIKK*, *17*, 101–116.

Furinghetti, F., & Morselli, F. (2004). Between Affect and Cognition: Proving at University Level. (2004). In Johnsen Hønes, M. & Berit Fuglestad, A (Eds), *Proceedings of the 28th*

Annual Conference of the International Group for the Psychology of Mathematics Education (Vol. 3, pp. 369-376). Bergen, Norway, Bergen University College.

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*, 71–90.

Gartmeier, M., Bauer, J., Fischer, M. R., Hoppe-Seyler, T., Karsten, G., Kiessling, C., . . . Prenzel, M. (2015). Fostering professional communication skills of future physicians and teachers: effects of e-learning with video cases and role-play. *Instructional Science*, *43*, 443–462. https://doi.org/10.1007/s11251-014-9341-6

Garuti, R., Boero, P., & Lemut, E. (1998). Cognitive unity of theorems and difficulty of proof. In A. Olivier & K. Newstead (Eds.), *Proceedings of the 22nd Conference of the International Group for the Psychology of Mathematics Education* (pp. 2–345).

Garuti, R., Boero, P., Lemut, E., & Mariotti, M. A. (1996). 'Challenging the traditional school approach to theorems: a hypothesis about the cognitive unity of theorems'. In L. Puig & A. Gutierrez (Eds.), *Proceedings of the 20th Annual Conference of the International Group for the Psychology of Mathematics Education* (Vol. 2, pp. 2-345). Spain: Universitat de València.

Gholamazad, S., Liljedahl, P., & Zazkis, R. (2003). One line proof: what can go wrong? In N. A. Pateman & Dougherty, B. J., Zilliox, J. T. (Eds.), *Proceedings of the 27th Annual Conference of the International Group for the Psychology of Mathematics Education Proceedings and PMENA* (Vol. 2, pp. 429–435). Honolulu, HI: University of Hawai'i.

Gibson, D. (1998). Students' Use Diagrams to Develop Proofs in an Introductory Analysis Course. *Research in Collegiate Mathematics Education. III*, *7*, 284–306.

Gillies, R. M. (2004). The effects of cooperative learning on junior high school students during small group learning. *Learning and Instruction*, *14*, 197–213.

Goodwin, C. (1994). Professional vision. *American Anthropologist*, *96*, 606–633.

Goos, M., Galbraith, P., & Renshaw, P. (2002). Socially mediated metacognition: Creating collaborative zones of proximal development in small group problem solving. *Educational Studies in Mathematics*, *49*, 193–223.

Graham, S., Weingart, S., & Milligan, I. (2012). Getting started with topic modeling and MALLET. Retrieved from https://programminghistorian.org/en/lessons/topic-modeling-and-mallet [10.11.2018].

Grenier, A. (2013). Research situations to learn logic and various types of mathematics reasoning and proofs. In B. Ubuz, Ç. Haser, & M. A. Mariotti (Eds.), *Proceedings of the 8th Congress of the European Society for Research in Mathematics Education* (pp. 136–145). Ankara, Turkey: Middle East Technical University.

Griffiths, P. A. (2000). Mathematics at the turn of the millennium. *The American Mathematical Monthly*, *107*, 1–14.

Gruenwald, N., Klymchuk, S., & Jovanoski, Z. (2004). Reducing the gap between the school and university mathematics: university lecturers' perspective. *New Zealand Mathematics Magazine, 41*(3), 12–24.

Gueudet, G. (2008). Investigating the secondary–tertiary transition. *Educational Studies in Mathematics*, *67*, 237–254.

Guzmán, M. de, Hodgson, B. R., Robert, A., & Villani, V. (1998). Difficulties in the passage from secondary to tertiary education, *Proceedings of the International Congress of Mathematicians,* Berlin, Documenta mathematica, extra volume ICM 1998, 747–762.

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

Hanna, G. (1995). Challenges to the importance of proof. *For the Learning of Mathematics*, *15*, 42–49.

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

Hanna, G., Jahnke, H. N., & Pulte, H. (2010). *Explanation and proof in mathematics*: *Philosophical and Educational Perspectives.* New York: Springer.

Hanna, G., Villiers, M. de, & International Program Committee. (2008). ICMI Study 19: Proof and proving in mathematics education. *ZDM*, *40*, 329–336.

Harel, G., & Sowder, L. (1998). Students’ proof schemes: Results from exploratory studies. *American Mathematical Society*, *7*, 234–283.

Harel, G., & Sowder, L. (2007). Toward Comprehensive Perspectives on the Learning and Teaching of Proof. In F. Lester (Ed.), *Second Handbook of Research on Mathematics Teaching and Learning* (Vol. 2, pp. 805-842). National Council of Teachers of Mathematics.

Healy, L., & Hoyles, C. (1998). *Justifying and proving in school mathematics: Technical report on the nationwide survey*: Institute of Education, University of London.

Healy, L., & Hoyles, C. (2000). A study of proof conceptions in algebra. *Journal for Research in Mathematics Education*, 396–428. https://doi.org/10.2307/749651

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). New York: Springer.

Heinze, A., Cheng, Y.-H., Ufer, S., Lin, F.-L., & Reiss, K. (2008a). Strategies to foster students' competencies in constructing multi-steps geometric proofs: Teaching experiments in Taiwan and Germany. *ZDM*, *40*, 443–453. https://doi.org/10.1007/s11858-008-0092-1

Heinze, A., Cheng, Y.-H., Ufer, S., Lin, F.-L., & Reiss, K. (2008b). Strategies to foster students' competencies in constructing multi-steps geometric proofs: Teaching experiments in Taiwan and Germany. *Zentralblatt für Didaktik der Mathematik*, *40*, 443–453.

Heinze, A., & Reiss, K. (2003). Reasoning and proof: Methodological knowledge as a component of proof competence. In M. A. Mariotti (Ed.), *International Newsletter of Proof,* No. 4-6/2003.

Heinze, A., & Reiss, K. (2004). The teaching of proof at the lower secondary level - a video study. *Zentralblatt für Didaktik der Mathematik*, *36*, 98–104.

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

Herlina, E., & Batusangkar, S. (2015). Advanced Mathematical Thinking and the Way to Enhance It. *Journal of Education and Practice*, *6*, 79–88.

Herweg, C., Seidel, T., & Dalehefte, I. M. (2005). Coding manual – Clear and structured teaching. In T. Seidel, M. Prenzel, & M. Kobarg (Eds.), *How to run a video study. Technical report of the IPN Video Study* (pp. 145–164). Münster: Waxmann.

Heublein, U. (2014). Student Drop-out from G erman Higher Education Institutions. *European Journal of Education*, *49*, 497–513.

Hiebert, J., Gallimore, R., Garnier, H., Bogard Givvin, K., Hollingsworth, H., Jacobs, J., Chui, A. M. Y., Wearne, D., Smith, M., Kersting, N., Manaster, A., Tseng, E., Etterbeek, W. Manaster, C., Gonzales, P. & Stigler, J. (2003). *Teaching Mathematics in Seven*

Countries: Results from the TIMSS 1999 Video Study (No. NCES 2003-013). Washington DC: U.S. Department of Education, National Center for Education Statistics.

Hinsz, V. B., Tindale, R. S., & Vollrath, D. A. (1997). The emerging conceptualization of groups as information processors. *Psychological Bulletin*, *121*, 43-64.

Hornikx, J., & Hahn, U. (2012). Reasoning and argumentation: Towards an integrated psychology of argumentation. *Thinking & Reasoning*, *18*, 225–243.

Hoyles, C., Newman, K., & Noss, R. (2001). Changing patterns of transition from school to university mathematics. *International Journal of Mathematical Education in Science and Technology*, *32*, 829–845.

Huda, N. (2016). University Students' Metacognitive Failures in Mathematical Proving Investigated Based on the Framework of Assimilation and Accommodation. *Educational Research and Reviews*, *11*, 1119–1128.

Inglis, M., & Aberdein, A. (2014). Beauty is not simplicity: an analysis of mathematicians' proof appraisals. *Philosophia Mathematica*, *23*, 87–109. https://doi.org/10.1093/philmat/nku014

Inglis, M., & Foster, C. (2018). Five decades of mathematics education research. *Journal for Research in Mathematics Education*, *49*, 462–500. https://doi.org/10.5951/jresematheduc.49.4.0462

Inglis, M., Mejia-Ramos, J. P., & Simpson, A. (2007). Modelling mathematical argumentation: The importance of qualification. *Educational Studies in Mathematics*, *66*, 3–21.

Järvelä, S., & Hadwin, A. F. (2013). New frontiers: Regulating learning in CSCL. *Educational Psychologist*, *48*, 25–39. https://doi.org/10.1080/00461520.2012.748006

Jiménez-Aleixandre, M. P., Bugallo Rodríguez, A., & Duschl, R. A. (2000). “Doing the lesson” or “doing science”: Argument in high school genetics. *Science Education*, *84*, 757–792.

Johnson, D. W., Johnson, R. T., & Smith, K. (2007). The state of cooperative learning in postsecondary and professional settings. *Educational Psychology Review*, *19*, 15–29.

Jordan, A., Krauss, S., Löwen, K., Blum, W., Neubrand, M., Brunner, M., . . . Baumert, J. (2008). Aufgaben im COACTIV-Projekt: Zeugnisse des kognitiven Aktivierungspotentials im deutschen Mathematikunterricht. *Journal Für Mathematik-Didaktik*, *29*, 83–107.

Kaartinen, S., & Kumpulainen, K. (2002). Collaborative inquiry and the construction of explanations in the learning of science. *Learning and Instruction*, *12*, 189–212.

Kahn, P. E., & Hoyles, C. (1997). The changing undergraduate experience: a case study of single honours mathematics in England and Wales. *Studies in Higher Education*, *22*, 349–362.

Kidron, I., & Dreyfus, T. (2014). Proof image. *Educational Studies in Mathematics*, *87*, 297–321.

Kneser, C., & Ploetzner, R. (2001). Collaboration on the basis of complementary domain knowledge: Observed dialogue structures and their relation to learning success. *Learning and Instruction*, *11*, 53–83.

Knipping, C. (2008). A method for revealing structures of argumentations in classroom proving processes. *ZDM*, *40*, 427. https://doi.org/10.1007/s11858-008-0095-y

Koedinger, K. R. (1998). Conjecturing and argumentation in high-school geometry students. *Designing Learning Environments for Developing Understanding of Geometry and Space*, 319–347.

Koeppen, K., Hartig, J., Klieme, E., & Leutner, D. (2008). Current issues in competence modeling and assessment. *Zeitschrift Für Psychologie/Journal of Psychology*, *216*, 61–73. https://doi.org/10.1027/0044-3409.216.2.61

Kollar, I., Fischer, F., & Slotta, J. D. (2007). Internal and external scripts in computer-supported collaborative inquiry learning. *Learning and Instruction*, *17*, 708–721. https://doi.org/10.1016/j.learninstruc.2007.09.021

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. https://doi.org/10.1016/j.learninstruc.2014.01.003

Komatsu, K. (2011). How do students generalize a conjecture through proving?: The importance of boundary cases between example and counterexample. In B. Ubuz (Ed.), *Proceedings of the 35th Conference of the International Group for the Psychology of Mathematics Education* (Vol. 3, pp. 89–96). Ankara: PME.

Komatsu, K. (2016). A framework for proofs and refutations in school mathematics: Increasing content by deductive guessing. *Educational Studies in Mathematics*, *92*, 147–162.

Komatsu, K., Tsujiyama, Y., & Sakamaki, A. (2014). Rethinking the discovery function of proof within the context of proofs and refutations. *International Journal of Mathematical*

Education in Science and Technology, *45*, 1053–1067. https://doi.org/10.1080/0020739X.2014.902135

Kopp, B., & Mandl, H. (2011). Fostering argument justification using collaboration scripts and content schemes. *Learning and Instruction*, *21*, 636–649.

Kosiol, T., Rach, S., & Ufer, S. (Which) Mathematics Interest is Important for a Successful Transition to a University Study Program? *International Journal of Science and Mathematics Education*, 1–22.

Krummheuer, G. (1995). The ethnography of argumentation. In P. Cobb & H. Bauersfeld (Eds.), *The emergence of mathematcial meaning: interaction in classroom cultures* (pp. 229–269). Hillsdale, NJ: Erlbaum.

Küchemann, D., & Hoyles, C. (2006). Secondary school pupils' approaches to proof-related tasks in geometry. In Hewitt, D. (Ed.), *Proceedings of the British Society for Research into Learning Mathematics* (Vol. *26*, pp. 53–58).

Lakatos, I. (1976). *Proofs and refutations: The logic of mathematical discovery*: Cambridge university press.

Langer, I., & Schulz von Thun, F. (2007). Messung komplexer Merkmale in Psychologie und Pädagogik: Ratingverfahren. In D. H. Rost (Ed.), *Standardwerke aus Psychologie und Pädagogik - Reprints* (4th ed.). Münster: Waxmann.

LeBreton, J. M., & Senter, J. L. (2008). Answers to 20 questions about interrater reliability and interrater agreement. *Organizational Research Methods*, *11*, 815–852. https://doi.org/10.1177/1094428106296642

Lee, E. Y. C., Chan, C. K. K., & van Aalst, J. (2006). Students assessing their own collaborative knowledge building. *International Journal of Computer-Supported Collaborative Learning*, *1*, 57–87.

Lee, K. (2016). Students' proof schemes for mathematical proving and disproving of propositions. *The Journal of Mathematical Behavior*, *41*, 26–44. https://doi.org/10.1016/j.jmathb.2015.11.005

Lee, V. E. (2000). Using hierarchical linear modeling to study social contexts: The case of school effects. *Educational Psychologist*, *35*, 125–141. https://doi.org/10.1207/S15326985EP3502_6

Leitão, S. (2000). The potential of argument in knowledge building. *Human Development*, *43*, 332–360. https://doi.org/10.1159/000022695

Leuders, T., & Philipp, K. (2013). Preparing students for discovery learning – skills for exploring mathematical patterns. In A. Lindmeier & A. Heinze (Eds.), *Proceedings of the 37th Conference of the International Group for the Psychology of Mathematics Education* (Vol. 3, pp. 241–248). Kiel, Germany: IPN.

Lin, F. L., Yang, K. L., Lee, K. H., Tabach, M., & Stylianides G. (2012). Task designing for conjecturing and proving: Developing principles based on practical tasks. In M. de Villiers & G. Hanna (Eds.), *Proof and Proving in Mathematics Education. The 19th ICMI Study* (pp. 305–323). Springer.

Lin, F.-L., Yang, K.-L., & Chen, C.-Y. (2004). The features and relationships of reasoning, proving and understanding proof in number patterns. *International Journal of Science and Mathematics Education*, *2*, 227–256.

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*, 165–196.

MacKinnon, D. P., Fairchild, A. J., & Fritz, M. S. (2007). Mediation analysis. *Annu. Rev. Psychol.*, *58*, 593–614.

Manin, Y. I. (1977). *A course in mathematical logic for mathematicians*. New York: Springer-Verlag.

Mariotti, M. A. (2006). Proof and proving in mathematics education. *Handbook of Research on the Psychology of Mathematics Education: Past, Present and Future*, 173–204.

Marston, C., & King, E. (2006). Factors that shape young people's sexual behaviour: a systematic review. *The Lancet*, *368*, 1581–1586.

Martin, W. G., & Harel, G. (1989). Proof frames of preservice elementary teachers. *Journal for Research in Mathematics Education*, 41–51.

Martinez, M. V., Brizuela, B. M., & Superfine, A. C. (2011). Integrating algebra and proof in high school mathematics: An exploratory study. *The Journal of Mathematical Behavior*, *30*, 30–47.

Mayweg-Paus, E., Thiebach, M., & Jucks, R. (2016). Let me critically question this! – Insights from a training study on the role of questioning on argumentative discourse. *International Journal of Educational Research*, *79*, 195–210. https://doi.org/10.1016/j.ijer.2016.05.017

McCallum, A. K. (2002). Mallet: A machine learning for language toolkit. University of Massachusetts, Amherst. http://mallet.cs.umass.edu/ [20.11.2018].

Meier, A., Spada, H., & Rummel, N. (2007). A rating scheme for assessing the quality of computer-supported collaboration processes. *International Journal of Computer-Supported Collaborative Learning*, *2*, 63–86. https://doi.org/10.1007/s11412-006-9005-x

Mejía-Ramos, J. P., & Inglis, M. (2009). 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).

Mercer, N., Dawes, L., Wegerif, R., & Sams, C. (2004). Reasoning as a scientist: Ways of helping children to use language to learn science. *British Educational Research Journal*, *30*, 359–377.

Miller, D., Infante, N., & Weber, K. (2018). How mathematicians assign points to student proofs. *The Journal of Mathematical Behavior*, *49*, 24–34.

Mills, M. (2014). A framework for example usage in proof presentations. *The Journal of Mathematical Behavior*, *33*, 106–118. https://doi.org/10.1016/j.jmathb.2013.11.001

Moore, R. C. (1994). Making the transition to formal proof. *Educational Studies in Mathematics*, *27*, 249–266.

Mueller, M., Yankelewitz, D., & Maher, C. (2012). A framework for analyzing the collaborative construction of arguments and its interplay with agency. *Educational Studies in Mathematics*, *80*, 369–387.

Muthén, L. K., & Muthén, B. O. (2015). *Mplus: Statistical analysis with latent variables: User's guide*. Muthén & Muthén Los Angeles.

Nadolski, R. J., Kirschner, P. A., & van Merriënboer, J. J. G. (2006). Process support in learning tasks for acquiring complex cognitive skills in the domain of law. *Learning and Instruction*, *16*, 266–278. https://doi.org/10.1016/j.learninstruc.2006.03.004

Nardi, E. (1996). *The Novice mathematician's encounter with mathematical abstraction: Tensions in concept-image construction and formulisation*. Oxford: University Oxford.

Newble, D. (2004). Techniques for measuring clinical competence: objective structured clinical examinations. *Medical Education*, *38*, 199–203.

Ohlsson, S., & Rees, E. (1991). The function of conceptual understanding in the learning of arithmetic procedures. *Cognition and Instruction*, *8*, 103–179. https://doi.org/10.1207/s1532690xci0802_1

Osborne, J. (2010). Arguing to learn in science: The role of collaborative, critical discourse. *Science*, *328*, 463–466. https://doi.org/10.1126/science.1183944

Otten, S., Bleiler-Baxter, S. K., & Engledowl, C. (2017). Authority and whole-class proving in high school geometry: The case of Ms. Finley. *The Journal of Mathematical Behavior*, *46*, 112–127. https://doi.org/10.1016/j.jmathb.2017.04.002

Ottinger, S., Kollar, I., & Ufer, S. (2016). Content and Form - All the Same or Different Qualities of Mathematical Arguments. In C. Csikos, A. Rausch, & J. Szitanyi (Eds.), *Proceedings of the 40th Conference of the International Group for the Psychology of Mathematics Education* (Vol. 4, pp. 19–26). Szeged, Hungary: PME.

Ozgur, Z., Ellis, A. B., Vinsonhaler, R., Dogan, M. F., & Knuth, E. (2017). From examples to proof: Purposes, strategies, and affordances of example use. *The Journal of Mathematical Behavior,* (in press).

Pastre, D. (1989). MUSCADET: an automatic theorem proving system using knowledge and metaknowledge in mathematics. *Artificial Intelligence*, *38*, 257–318. https://doi.org/10.1016/0004-3702(89)90035-0

Pease, A., & Martin, U. (2012). Seventy four minutes of mathematics: An analysis of the third Mini-Polymath project. In AISB/IACAP World Congress 2012 (Ed.), *Proceedings of AISB/IACAP 2012* (pp. 52–105).

Pease, A., Smaill, A., Colton, S., & Lee, J. (2009). Bridging the gap between argumentation theory and the philosophy of mathematics. *Foundations of Science*, *14*, 111–135.

Pedemonte, B. (2007). How can the relationship between argumentation and proof be analysed? *Educational Studies in Mathematics*, *66*, 23–41.

Pedemonte, B. (2008). Argumentation and algebraic proof. *Zentralblatt für Didaktik der Mathematik*, *40*, 385–400.

Pedemonte, B., & Buchbinder, O. (2011). Examining the role of examples in proving processes through a cognitive lens: the case of triangular numbers. *Zentralblatt für Didaktik der Mathematik*, *43*, 257–267.

Pedemonte, B., & Reid, D. (2011). The role of abduction in proving processes. *Educational Studies in Mathematics*, *76*, 281–303.

Perelman, C. (1979). The rational and the reasonable. In C. Perelman (Ed.), *The New Rhetoric and the Humanities* (pp. 117–123). Bosten: D. Reidel.

Philipp, K. (2012). *Experimentelles Denken: theoretische und empirische Konkretisierung einer mathematischen Kompetenz*. Springer-Verlag.

Pinto, M., & Tall, D. (2002). Building formal mathematics on visual imagery: A case study and a theory. *For the Learning of Mathematics*, *22*, 2–10.

Plaxco, D., & Wawro, M. (2015). Analyzing student understanding in linear algebra through mathematical activity. *The Journal of Mathematical Behavior*, *38*, 87–100.

Polya, G. (1945). *How to solve it*. New Jersey: Princeton University Press.

Praslon, F. (2000). Continuités et ruptures dans la transition Terminale S/DEUG Sciences en analyse: Le cas de la notion de dérivée et son environnement. *Publications Mathématiques Et Informatiques De Rennes*, *3*, 1–27.

Rach, S. (2014). *Charakteristika von Lehr-Lern-Prozessen im Mathematikstudium: Bedingungsfaktoren für den Studienerfolg im ersten Semester*: Waxmann Verlag.

Rach, S., & Heinze, A. (2011). Studying mathematics at the university: the influence of learning strategie. In B. Ubuz (Ed.), *Proceedings of the 35th Conference of the International Group for the Psychology of Mathematics Education* (pp. 9–16). Ankara: PME.

Raman, M. (2003). Key ideas: What are they and how can they help us understand how people view proof? *Educational Studies in Mathematics*, *52*, 319–325.

Reichersdorfer, E., Vogel, F., Fischer, F., Kollar, I., Reiss, K., & Ufer, S. (2012). Different collaborative learning settings to foster mathematical argumentation skills. In T. Tso (Ed.), *Proceedings of the 36th conference of the International Group for the Psychology of Mathematics Education* (pp. 345–352).

Reif, F. (2008). *Applying cognitive science to education: thinking and learning in scientific and other complex domains*: MIT press.

Reinholz, D. L. (2016). Improving calculus explanations through peer review. *The Journal of Mathematical Behavior*, *44*, 34–49. https://doi.org/10.1016/j.jmathb.2016.10.001

Roschelle, J., & Teasley, S. D. (1995). The construction of shared knowledge in collaborative problem solving. In C. E. O'Malley (Ed.), *Computer-Supported Collaborative Learning* (pp. 69–197). Berlin: Springer.

Roth, K. J. (2009). Using video studies to transform science teaching and learning: results from the STeLLA professional development program. In T. Janik, & T. Seidel (Eds.)*,The* *power of video studies in investigating teaching and learning in the classroom* (pp. 23-37). Münster, Germany: Waxmann.

Sampson, V., & Clark, D. (2009). The impact of collaboration on the outcomes of scientific argumentation. *Science Education*, *93*, 448–484.

Sandefur, J., Mason, J., Stylianides, G. J., & Watson, A. (2013). Generating and using examples in the proving process. *Educational Studies in Mathematics*, *83*, 323–340. https://doi.org/10.1007/s10649-013-9510-6

Sandoval, W. A. (2003). Conceptual and epistemic aspects of students' scientific explanations. *The Journal of the Learning Sciences*, *12*, 5–51. https://doi.org/10.1207/S15327809JLS1201_2

Savic, M. (2015a). The incubation effect: How mathematicians recover from proving impasses. *The Journal of Mathematical Behavior*, *39*, 67–78. https://doi.org/10.1016/j.jmathb.2015.06.001

Savic, M. (2015b). On similarities and differences between proving and problem solving. *Journal of Humanistic Mathematics*, *5*, 60–89. https://doi.org/10.5642/jhummath.201502.06

Schiefele, U., Streblow, L., & Brinkmann, J. (2007). Aussteigen oder Durchhalten. *Zeitschrift Für Entwicklungspsychologie Und Pädagogische Psychologie*, *39*, 127–140. https://doi.org/10.1026/0049-8637.39.3.127

Schoenfeld, A. (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: McMillan.

Schulmeiß, I., Seidel, T., & Meyer, L. (2003). Vermischung von Lern-und Leistungssituationen. In T. Seidel, M. Prenzel, R. Duit, M. Lehrke (Hrsg.), *Technischer Bericht zur Videostudie. Lehr-Lern-Prozesse im Physikunterricht* (S. 229-238). Kiel, Germany: IPN.

Schwaighofer, M., Vogel, F., Kollar, I., Ufer, S., Strohmaier, A., Terwedow, I., . . . Fischer, F. (2017). How to combine collaboration scripts and heuristic worked examples to foster mathematical argumentation – when working memory matters. *International Journal of Computer-Supported Collaborative Learning*, *12*, 281–305. https://doi.org/10.1007/s11412-017-9260-z

Schwartz, D. L. (1995). The emergence of abstract representations in dyad problem solving. *The Journal of the Learning Sciences*, *4*, 321–354.

Schwarz, B. B., Hershkowitz, R., & Prusak, N. (2010). Argumentation and mathematics. *Educational Dialogues: Understanding and Promoting Productive Interaction*, *115*, 141.

Seidel, T. (2005). Video analysis strategies of the IPN Video Study – a methodological overview. In T. Seidel, M. Prenzel, & M. Kobarg (Eds.), *How to run a video study. Technical report of the IPN Video Study* (pp. 70–78). Münster: Waxmann.

Seidel, T., Prenzel, M., & Kobarg, M. (Eds.). (2005). *How to run a video study. Technical report of the IPN Video Study*: Münster: Waxmann.

Selden, A., & Selden, J. (1987). Errors and misconceptions in college level theorem proving. In *Proceedings of the second international seminar on misconceptions and educational strategies in science and mathematics* (Vol. 3, pp. 457–470). New York: Cornell University.

Selden, A., & Selden, J. (2008). Overcoming students' difficulties in learning to understand and construct proofs. *Making the Connection: Research and Teaching in Undergraduate Mathematics*, 95–110.

Selden, A., & Selden, J. (2011). Mathematical and non-mathematical university students' proving difficulties. In L. R. Wiest & T. D. Lamberg (Eds.), *Proceedings of the 33rd annual conference of the North American chapter of the International Group for the Psychology of Mathematics Education* (pp. 675–683). Reno, NV.

Selden, A., & Selden, J. (2013a). The roles of behavioral schemas, persistence, and self-efficacy in proof construction. In B. Ubuz, C. Hasar, & M. A. Mariotti (Eds.), *Proceedings of the 8th Congress of the European Society for Research in Mathematics Education [CERME-8]* (pp. 246-255). Ankara, Turkey: Middle East Technical University.

Selden, A., & Selden, J. (2015). Validations of proofs as a type of reading and sense-making. In K. Beswick, T. Muir, & J. Wells (Eds.), *Proceedings of the 39th Annual Conference of the International Group for the Psychology of Mathematics Education* (Vol. 4, pp. 145–152). Hobart, Australia.

Selden, A., McKee, K., & Selden, J. (2010). Affect, behavioural schemas and the proving process. *International Journal of Mathematical Education in Science and Technology*, *41*, 199–215.

Selden, A., & Selden, J. (2009). Teaching proving by coordinating aspects of proofs with students' abilities. In D. A. Stylianou, M. L. Blanton, & E. J. Knuth (Eds.), *Teaching and learning across the grades: A K-16 perspective* (pp. 339–354). New York, NY: Routledge.

Selden, A., & Selden, J. (2013b). Proof and problem solving at university level. *The Mathematics Enthusiast*, *10*, 303–334.

Selden, A., Selden, J., & Benkhalti, A. (2018). Proof Frameworks: A Way to Get Started. *Problems, Resources, and Issues in Mathematics Undergraduate Studies*, *28*, 31–45.

Selden, J., Benkhalti, A., & Selden, A. (2014). An analysis of transition-to-proof course students' proof constructions with a view towards course redesign. In T. Fukawa-Connolly, G. Karakok, K. Keene, & M. Zandieh (Eds.), *Proceedings of the 17th Annual Conference on Research in Undergraduate Mathematics Education* (pp. 246–259).

Selden, J., & Selden, A. (2015). A perspective for university students' proof construction. In *Proceedings of the 18th Annual Conference on Research in Mathematics Education* (pp. 45–59).

Selden, J., & Selden, A. (1995). Unpacking the logic of mathematical statements. *Educational Studies in Mathematics*, *29*, 123–151.

Shernoff, E. S., & Kratochwill, T. R. (2003). The Application of behavioral assessment methodologies in educational settings. In S. N. Haynes & E. H. Heiby (Eds.), *Comprehensive handbook of psychological assessment* (pp.365–385). New York: John Wiley.

Sommerhoff, D. (2017). *The individual cognitive resources underlying students' mathematical argumentation and proof skills* (Doctoral dissertation). München: LMU München. Retrieved from: https://edoc.ub.uni-muenchen.de/22687/1/Sommerhoff_Daniel.pdf [18.11.2018].

Sommerhoff, D., Ufer, S., & Kollar, I. (2016). The impact of cognitive resources on students' performance in mathematical proof construction: 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 (ICLS)* (pp. 811–814). International Society of the Learning Sciences.

Spencer, L. M., & Spencer, P. S. M. (2008). *Competence at Work models for superior performance*: John Wiley & Sons.

Stahl, G. (2010). Group cognition as a foundation for the new science of learning. In M. S. Khine & I. M. Saleh (Eds.), *New science of learning: Cognition, computers and collaboration in education* (pp. 23-44). New York, NY: Springer.

Staples, M. (2007). Supporting whole-class collaborative inquiry in a secondary mathematics classroom. *Cognition and Instruction*, *25*, 161–217.

Steyvers, M., & Griffiths, T. (2007). Probabilistic topic models. In T. Landauer, D McNamara, S. Dennis, & W. Kintsch, *Handbook of Latent Semantic Analysis* (Vol. *427*, pp. 424–440).

Stylianides, A. J. (2007). Proof and proving in school mathematics. *Journal for Research in Mathematics Education*, 289–321. https://doi.org/10.2307/30034869

Stylianides, A. J., Bieda, K. N., & Morselli, F. (2016). Proof and argumentation in mathematics education research. In A. Gutierez, G. Leder, & P. Boero (Eds.), *2nd handbook on the psychology of mathematics education* (pp. 315–351). Rotterdam: Sense Publishers.

Stylianides, G. J. (2010). Engaging Secondary Students in Reasoning and Proving. *Mathematics Teaching*, *219*, 39–44.

Stylianides, G. J., Sandefur, J., & Watson, A. (2016). Conditions for proving by mathematical induction to be explanatory. *The Journal of Mathematical Behavior*, *43*, 20–34.

Stylianides, G. J., & Stylianides, A. J. (2009). Facilitating the transition from empirical arguments to proof. *Journal for Research in Mathematics Education*, 314–352.

Stylianides, G. J., Stylianides, A., & Weber, K. (2017). Research on the teaching and learning of proof: Taking stock and moving forward. In J. Cai (Ed.), *Compendium for Research in Mathematics Education* (pp. 237–266). Reston, VC: National Council of Teachers of Mathematics.

Stylianou, D. A., Blanton, M. L., & Knuth, E. J. (2010). *Teaching and learning proof across the grades: A K-16 perspective*: Routledge.

Tall, D. (1991). *Advanced mathematical thinking*: Springer Science & Business Media.

Teasley, S. D. (1997). Talking about reasoning: How important is the peer in peer collaboration? In L. B. Resnick & R. Säljö & C. Pontecorvo & B. Burge (Eds.), *Discourse, tools and reasoning: Essays on situated cognition* (pp. 361384). Berlin: Springer.

Thomas, J., & Harden, A. (2008). Methods for the thematic synthesis of qualitative research in systematic reviews. *BMC Medical Research Methodology*, *8*, 45.

Thomas, M. O. J., & Klymchuk, S. (2012). The school–tertiary interface in mathematics: teaching style and assessment practice. *Mathematics Education Research Journal*, *24*, 283–300.

Thurston, W. P. (1998). On proof and progress in mathematics. *New Directions in the Philosophy of Mathematics*, 337–355.

Toulmin, S. (1958). *The uses of argument*: Cambridge: Cambridge University Press.

Tristanti, L. B., Sutawidjaja, A., As'ari, A. R., & Muksar, M. (2016). The construction of deductive warrant derived from inductive warrant in preservice-teacher mathematical argumentations. *Educational Research and Reviews*, *11*, 1696–1708.

Uebersax, J. (2010). Statistical methods for rater and diagnostic agreement. Retrieved from: *http://www. john-uebersax.com/stat/agree.htm* [09.10.2018].

Ufer, S., Heinze, A., & Reiss, K. (2008). Individual predictors of geometrical proof competence. In O. Figueras, J. L. Cortina, S. Alatorre, T. Rojano, & A. Sepulveda (Eds.), *Proceedings of the Joint Meeting of PME* (Vol. 4, pp. 361–368).

Ufer, S., Heinze, A., & Reiss, K. (2009). 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).

Uğurel, I., Moralı, S., Koyunkaya, M. Y., & Karahan, Ö. (2016). Pre-service Secondary Mathematics Teachers' Behaviors in the Proving Process. *Eurasia Journal of Mathematics, Science & Technology Education*, *12*(2), 203-301.

Van Boxtel, C., van der Linden, J., & Kanselaar, G. (2000). Collaborative learning tasks and the elaboration of conceptual knowledge. *Learning and Instruction*, *10*, 311–330.

Van Es, E. A., & Sherin, M. G. (2002). Learning to notice: Scaffolding new teachers' interpretations of classroom interactions. *Journal of Technology and Teacher Education*, *10*, 571–596.

VanSpronsen, H. D. (2008). *Proof processes of novice mathematics proof writers*. (Doctoral dissertation). Retrieved from: https://scholarworks.umt.edu/etd/778/ [26.10.2018].

Vidakovic, D., & Martin, W. O. (2004). Small-group searches for mathematical proofs and individual reconstructions of mathematical concepts. *The Journal of Mathematical Behavior*, *23*, 465–492. https://doi.org/10.1016/j.jmathb.2004.09.006

Villiers, M. de. (1999). *Rethinking proof with the Geometer's Sketchpad.* Key Curriculum Press, Emeryville, CA.

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*, 477–500. https://doi.org/10.1007/s11251-016-9380-2

Vogel, F., Wecker, C., Kollar, I., & Fischer, F. (2017). Socio-cognitive scaffolding with computer-supported collaboration scripts: A meta-analysis. *Educational Psychology Review*, *29*, 477–511.

Vollstedt, M., Heinze, A., Gojdka, K., & Rach, S. (2014). A Framework for Examining the Transformation of Mathematics and Mathematics Learning in the Transition from School to University: An Analysis of German Textbooks from Upper Secondary School and the First Semester. In S. Rezat, M. Hattermann & A. Peter-Koop (Hrsg.), *Transformation – A Fundamental Idea of Mathematics Education* (pp. 29–50). New York: Springer.

Webb, N. M. (1982). Group composition, group interaction, and achievement in cooperative small groups. *Journal of Educational Psychology*, *74*, 475.

Webb, N. M. (1989). Peer interaction and learning in small groups. *International Journal of Educational Research*, *13*, 21–39.

Webb, N. M., Nemer, K. M., & Zuniga, S. (2002). Short circuits or superconductors? Effects of group composition on high-achieving students' science assessment performance. *American Educational Research Journal*, *39*, 943–989. https://doi.org/10.3102/00028312039004943

Weber, K. (2001). Student difficulty in constructing proofs: The need for strategic knowledge. *Educational Studies in Mathematics*, *48*, 101–119.

Weber, K. (2004). A Framework for Describing the Processes that Undergraduates Use to Construct Proofs. In M. J. Hoines & 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: 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*, 351–360.

Weber, K. (2009). How syntactic reasoners can develop understanding, evaluate conjectures, and generate counterexamples in advanced mathematics. *The Journal of Mathematical Behavior*, *28*, 200–208. https://doi.org/10.1016/j.jmathb.2009.08.001

Weber, K., & Alcock, L. (2004). Semantic and syntactic proof productions. *Educational Studies in Mathematics*, *56*, 209–234.

Weber, K., Fukawa-Connelly, T. P., Mejía-Ramos, J. P., & Lew, K. (2016). How to Help Students Understand Lectures in Advanced Mathematics. *Notices of the AMs*, *63*.

Weber, K., Inglis, M., & Mejia-Ramos, J. P. (2014). How mathematicians obtain conviction: Implications for mathematics instruction and research on epistemic cognition. *Educational Psychologist*, *49*, 36–58. https://doi.org/10.1080/00461520.2013.865527

Wecker, C., & Fischer, F. (2014). From guided to self-regulated performance of domain-general skills: The role of peer monitoring during the fading of instructional scripts. In T. Seidel & A. Krapp (Eds.), *Pädagogische Psychologie* (pp. 746–756). Weinheim: Beltz.

Wegerif, R. (2008). Dialogic or dialectic? The significance of ontological assumptions in research on educational dialogue. *British Educational Research Journal*, *34*, 347–361.

Weinberger, A., & Fischer, F. (2006). A framework to analyze argumentative knowledge construction in computer-supported collaborative learning. *Computers & Education*, *46*, 71–95. https://doi.org/10.1016/j.compedu.2005.04.003

Weinert, F. E. (2001). Concept of competence: A conceptual clarification. Concept of competence: A conceptual clarification. In D. S. Rychen & L. H. Salganik (Eds.), *Defining and selecting key competencies* (pp. 45-65). Ashland, OH, US: Hogrefe & Huber Publishers.

Wiesbeck, A. B. (2015). *An evaluation of simulated conversations as an assessment of pre-service teachers' communication competence in parent-teacher conversations* (Doctoral dissertation). München: TU München. Retrieved from: https://mediatum.ub.tum.de/?id=1271404 [20.11.2018].

Winkel, B. (2015). Informed conjecturing of solutions for differential equations in a modeling context. *Problems, Resources, and Issues in Mathematics Undergraduate Studies*, *25*, 158–169.

Wirtz, M. A., & Caspar, F. (2002). *Beurteilerübereinstimmung und Beurteilerreliabilität: Methoden zur Bestimmung und Verbesserung der Zuverlässigkeit von Einschätzungen mittels Kategoriensystemen und Ratingskalen*: Hogrefe.

Wolf, E. J., Harrington, K. M., Clark, S. L., & Miller, M. W. (2013). Sample size requirements for structural equation models: An evaluation of power, bias, and solution propriety. *Educational and Psychological Measurement*, *73*, 913–934.

Wood, D. (1999). Teaching the young child: Some relationships between social interaction, language and thought. *Lev Vygotsky: Critical Assessments*, *3*, 259–275.

Yackel, E. (2001). Explanation, Justification and Argumentation in Mathematics Classrooms. In M. Van den Heuvel-Panhuizen (Ed.), *Proceedings of the 25th conference of the* *international group for the psychology of mathematics education* (Vol. 1, pp. 1–9). Utrecht, Olanda: PME.

Yackel, E., & Cobb, P. (1996). Sociomathematical norms, argumentation, and autonomy in mathematics. *Journal for Research in Mathematics Education*, 458–477.

Yang, K.-L. (2012). Providing opportunities for students to create mathematics. *Procedia-Social and Behavioral Sciences*, *46*, 3905–3909. https://doi.org/10.1016/j.sbspro.2012.06.170

Yetter, G., Gutkin, T. B., Saunders, A., Galloway, A. M., Sobansky, R. R., & Song, S. Y. (2006). Unstructured collaboration versus individual practice for complex problem solving: A cautionary tale. *The Journal of Experimental Education*, *74*, 137–160.

Yopp, D. A., & Ely, R. (2016). When does an argument use a generic example? *Educational Studies in Mathematics*, *91*, 37–53.

Zandieh, M., & Rasmussen, C. (2010). Defining as a mathematical activity: A framework for characterizing progress from informal to more formal ways of reasoning. *The Journal of Mathematical Behavior*, *29*, 57–75. https://doi.org/10.1016/j.jmathb.2010.01.001

Zandieh, M., Roh, K. H., & Knapp, J. (2014). Conceptual blending: Student reasoning when proving “conditional implies conditional” statements. *The Journal of Mathematical Behavior*, *33*, 209–229.

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). Netherlands: Springer.

Zazkis, D., & Villanueva, M. (2016). Student conceptions of what it means to base a proof on an informal argument. *International Journal of Research in Undergraduate Mathematics Education*, *2*, 318–337.

Zazkis, D., Weber, K., & Mejía-Ramos, J. P. (2015). Two proving strategies of highly successful mathematics majors. *The Journal of Mathematical Behavior*, *39*, 11–27. https://doi.org/10.1016/j.jmathb.2015.04.003

Zazkis, D., Weber, K., & Mejía-Ramos, J. P. (2016). Bridging the gap between graphical arguments and verbal-symbolic proofs in a real analysis context. *Educational Studies in Mathematics*, *93*, 155–173. https://doi.org/10.1007/s10649-016-9698-3

Eidesstattliche Versicherung

(Siehe Promotionsordnung vom 12.07.11, § 8, Abs. 2 Pkt. .5.)

Hiermit erkläre ich an Eidesstatt, dass die Dissertation von mir selbstständig, ohne unerlaubte Beihilfe angefertigt ist.

Ottinger, Sarah

Name, Vorname

München, 22.11.2018 Sarah Ottinger

Ort, Datum Unterschrift Doktorand/in