When young students fail to productively learn with Productive Failure – analyzing core learning mechanisms

Inauguraldissertation zur Erlangung des akademischen Grades
einer Doktorin der Philosophie

vorgelegt von
Claudia Mazziotti
aus Lippstadt

Bochum im April 2017

Referentin: Prof. Dr. Nikol Rummel
Korreferent: Prof. Dr. Vincent Aleven
Dekanin: Prof. Dr. Corinna Mieth

Datum der Disputation: 13. Juli 2017
Acknowledgements

As without support this thesis would not have been possible, I would like to thank a number of persons who have helped me at various stages of this work.

My deepest gratitude goes first to my advisor Prof. Dr. Nikol Rummel who always took care of me, believed in me, supported me and provided me with smaller and bigger challenges in my zone of proximal development. Without her, this thesis would not have been possible. Thanks to my co-advisor Prof. Dr. Vincent Aleven for his support and feedback on different aspects of my work in the last years. Special thanks also go to my former colleague Dr. Katharina Loibl and my collaboration partner Dr. Anne Deiglmayr for their huge amount of patience and persistence in helping to improve my thesis on various strands. Thank you Anne for the great research stay at the ETH Zurich. In addition, a thank you to Prof. Dr. Joachim Wirth for providing me with feedback during our joint research colloquia.

This thesis would not have been possible without the effort of the participating schools and the consent of the parents to the involvement of their children in my studies.

I am grateful for the support from all my current and former colleagues, namely Dr. Katharina Loibl, Dr. Michael Wiedmann, Dr. Alexandra Funk, Valentina Nachtigall, Julia Erdmann, Sebastian Strauss, Christian Hartmann, Nesrin Denizer, Dr. Astrid Wichmann, Dr. Julia Eberle, Dr. Malte Elson and Hamza Sati. In the last years, everyone has always been ready to help, to collaboratively search for solutions, to encourage me and to cheer me up. I especially thank Dr. Michael Wiedmann for having my back by supporting me in many non-thesis related issues and tasks. I would also like to thank our visiting researcher Dr. Anouschka van Leeuwen, Sebastian Strauss and Julia Erdmann who helped with the final steps of the thesis. Furthermore, I am thankful to all student assistants for their support: special thanks to Anna Müller and Gina Jochheim (former Master candidate) for their help in coding the process data, to Tim Schönfeld and Lena Hempert for their help with formatting issues, and to Gina Krüger who solved a lot of teaching related issues. I am very grateful for the support from all research interns, namely Jan Schröder, Gina Krüger, Nadja Röder, Tugce Durgut, Martin Koch und Sandra Schulz. In addition, I would like to thank Dr. Alice Hansen and Heike Murglat for their help in improving the learning and test material from a mathematics education point of view.
My appreciation also extends to my family and friends. Thanks to my parents who always believed in me and stood by my side. Thanks to my friends for their emotional support, for their understanding and for keep asking me to participate in free-time activities.

As no thesis can be written without funding, I would like to thank the European Union which partly funded me during the last years (i.e., Union Seventh Framework Programme, FP7/2007-2013, under grant agreement n° 318051 - iTalk2Learn project).
# Table of Contents

Acknowledgements ........................................................................................................ ii  
Table of Contents ........................................................................................................ iv  
List of Figures and Tables ............................................................................................ vi  
Zusammenfassung (German summary) ........................................................................ vii  

1 Introduction .............................................................................................................. 1  

2 Theoretical background ......................................................................................... 6  
  2.1 Productive Failure .............................................................................................. 6  
  2.2 Delaying instruction for young learners ............................................................ 8  
  2.3 Form of instruction ............................................................................................ 10  
  2.4 Social form of learning .................................................................................... 12  
  2.5 Analyses of learning mechanisms underlying Productive Failure .................... 14  
  2.5.1 Students’ prior knowledge activation and differentiation ............................ 15  
  2.5.2 Students’ awareness of knowledge gaps ....................................................... 16  
  2.5.3 Students’ recognition of deep features of the target domain ...................... 17  
  2.6 Summary of hypotheses .................................................................................... 18

3 Study 1 ....................................................................................................................... 21  
  3.1 Method .............................................................................................................. 21  
  3.1.1 Design .......................................................................................................... 21  
  3.1.2 Sample .......................................................................................................... 22  
  3.1.3 Learning material .......................................................................................... 22  
  3.1.4 Instruments ................................................................................................... 32  
  3.1.5 Procedure ...................................................................................................... 33  
  3.2 Results and discussion ..................................................................................... 34  
  3.2.1 Findings and implications concerning study goal 1 and 2 ............................ 34  
  3.2.2 Preliminary results and discussion concerning the main hypotheses (study goal 3) ................................................................................................................. 39

4 Study 2 ....................................................................................................................... 42  
  4.1 Method .............................................................................................................. 43  
  4.1.1 Design .......................................................................................................... 44  
  4.1.2 Sample .......................................................................................................... 45  
  4.1.3 Posttest ......................................................................................................... 45  
  4.1.4 Process analyses to detect underlying learning mechanisms ...................... 46  
  4.2 Results .............................................................................................................. 57  
  4.2.1 Prior analyses ............................................................................................... 57
4.2.2 Conceptual knowledge posttest score ........................................................... 57
4.2.3 Students’ solutions and conceptual knowledge ............................................ 58
4.2.4 Types of collaborative utterances, students’ solutions and conceptual knowledge at posttest ................................................................. 60
4.3 Discussion ..................................................................................................... 62
4.3.1 Students’ prior knowledge activation and differentiation ............................ 63
4.3.2 Students’ awareness of knowledge gaps ...................................................... 65
4.3.3 Students’ recognition of deep features of the target domain ................. 68

5 General discussion and conclusion ................................................................... 70
5.1 Discussion of results ..................................................................................... 73
5.2 Limitations and outlook .............................................................................. 77
5.2.1 Analyzing collaborative learning ............................................................... 77
5.2.2 Further research on collaborative learning within PF ............................... 81
5.2.3 Additional factors underlying PF? .............................................................. 82
5.3 Pedagogical implications ............................................................................. 87
5.4 Conclusion ................................................................................................... 90

6 References ................................................................................................... 92

7 Appendixes ................................................................................................ 106
7.1 Appendix A: Additional information ......................................................... 106
7.1.1 Padberg’s equivalent fraction problem .................................................... 106
7.1.2 Role cards ............................................................................................... 107
7.2 Appendix B: Curriculum Vitae ................................................................. 108
List of Figures and Tables

Figure 3.1. Basic template of the equivalent fraction problem across experimental conditions ................................................................. 24
Figure 3.2. The equivalent fraction problem within the problem-solving phase of Productive Failure .......................................................... 26
Figure 3.3. The equivalent fraction problem within the problem-solving phase of Direct Instruction ......................................................... 27
Figure 3.4. Activities to practice the division-feature and the comparison-feature. 31
Figure 3.5. Example item of the conceptual knowledge posttest in Study 1 .......... 33
Figure 3.6. Example for a typical student solution of the PF-problem (plus translation). ........................................................................... 35
Figure 4.1. Example item of the conceptual knowledge posttest in Study 2 .......... 46

Table 2.1. Overview about PF-similar studies conducted with elementary school students ................................................................. 9
Table 3.1. Overview of experimental conditions and learning phases in Study 1... 22
Table 3.2. Comparing and contrasting activity in the instruction phase for problem-solving step 1 ................................................................. 29
Table 3.3. Mean and standard deviations of pretest scores of Study 1 ............... 40
Table 3.4. Means and standard deviations of students’ conceptual knowledge at posttest of Study 1 ............................................................. 40
Table 4.1. Overview about experimental conditions and the learning phases in Study 2 ............................................................................. 45
Table 4.2. Rating of the solution quality with examples .................................... 49
Table 4.3. Overview of coding scheme with coding examples .......................... 55
Table 4.4. Mean and standard deviations of pretest scores of Study 2 ............... 57
Table 4.5. Means and standard deviations of students’ conceptual posttest scores at posttest of Study 2 ......................................................... 58
Table 4.6. Number of student-generated solution ideas and solution quality of the PF-Coll and PF-Ind conditions ......................................... 59
Table 4.7. Means and standard deviations of students’ collaborative utterances in the PF-Coll condition, measured at the individual level (n=68) ............ 60
Zusammenfassung (German summary)


Während die Effektivität von PF für den konzeptuellen Wissenserwerb von SuS am Ende der Sekundarstufe 1 gut erforscht ist, ist bislang weniger über die Effektivität von PF für den konzeptuellen Wissenserwerb von jüngeren SuS am Ende der Grundschule oder zu Beginn der Sekundarstufe 1 bekannt. Es lassen sich Studien identifizieren, die zwar nicht PF als Lernansatz implementieren, aber einen ähnlichen Ansatz, bei dem SuS im Grundschulalter erst Aufgaben bearbeiten und dann Instruktion erhalten (Decaro & Rittle-Johnson, 2012; Loehr, Fyfe & Rittle-Johnson, 2014; Fyfe, Decaro & Rittle-Johnson, 2014; Malten & Klahr, 2012). Eine weitere Gemeinsamkeit zwischen dieser Gruppe von Studien und PF-Studien mit älteren SuS, ist der Vergleich zwischen SuS, die zuerst Aufgaben bearbeiten und dann Instruktion erhalten, mit SuS, die zuerst Instruktion erhalten und dann Aufgaben bearbeiten (Kontrollbedingung), sowie die Erfassung des konzeptuellen


Um die zuvor genannten Hypothesen und die Entfaltung der drei Lernmechanismen zu untersuchen, wurden zwei quasi-experimentelle Studien durchgeführt und Aufgabenbearbeiten vor Instruktion als PF-Lernansatz umgesetzt. Beide Studien variierten die Faktoren Zeitpunkt der (angepassten) Instruktion (Aufgabenbearbeiten vor Instruktion, PF vs. Aufgabenbearbeiten nach Instruktion, DI) und die Sozialform des Lernens während des Aufgabenbearbeitens (kooperatives Lernen während des Aufgabenbearbeitens vs. individuelles Lernen während des Aufgabenbearbeitens). Da für die junge Zielgruppe bislang noch kein PF-Material entwickelt wurde, zielte die erste Studie auf die 1) Entwicklung von Lern- und Testmaterial, 2) die Entwicklung von Analysemethoden zur Erfassung der Lernmechanismen sowie auf 3) die anfängliche Überprüfung der beiden Haupthypothesen (Hypothesen 1 und 2) ab. Hierfür wurden drei Bedingungen (kooperatives Aufgabenbearbeiten vor Instruktion, PF-Koop, vs. individuelles Aufgabenbearbeiten nach Instruktion, PF-Ind, vs. kooperatives Aufgabenbearbeiten nach Instruktion, DI-
Koop) implementiert und eine Stichprobe im Umfang von 52 SuS der vierten Grundschulklass e (in Deutschland) rekrutiert. Das Ziel der zweiten Studie lag darin, alle Hypothesen systematisch zu testen und die Entfaltung der Lernmechanismen mit Hilfe der in Studie 1 entwickelten Analysemethoden zu untersuchen. Hierfür wurde eine wesentlich größere Stichprobe (N= 228) rekrutiert und ein komplettes 2x2 Studiendesign realisiert.


Auch wenn sich auf der Ebene des konzeptuellen Wissens kein Vorteil des kooperativen Lernens gegenüber dem individuellen Lernen zeigte (vgl. Hypothese 2), deutet die Reihe an Kooperationsprozessanalysen auf einen Vorteil der Kooperation während des Lernens in der Aufgabenbearbeitungsphase hin. Es zeigte sich nämlich, dass kooperativ lernende SuS deskriptiv mehr Lösungsideen und auch qualitativ hochwertigere Schülerlösungen

1 Introduction

When and how to provide learners with assistance is a fundamental question in the fields of the learning sciences, educational psychology, and educational science (Koedinger & Aleven, 2007). Especially when students learn about a new topic the so called assistance dilemma (Koedinger & Aleven, 2007) has to some extent separated scientists and practitioners into two opposing camps (see also Tobias & Duffy, 2009): In the one camp instructivists advocate to provide novice learners with a high level of assistance (e.g., in form of direct feedback or instruction) right from the beginning of the learning process (cf. Kirschner, Sweller, & Clark, 2006; Klahr & Nigam, 2004). This way, the learning process of novice learners is not overshadowed by students’ high cognitive load (Sweller, 1988, 2009). In contrast, researchers from the constructivist camp advocate to enable learners to first solve problems on their own and to receive assistance (e.g., in form of an instruction) afterwards or to a much lower extent during problem solving (cf. Hmelo-Silver, Duncan, & Chinn, 2007; Kapur, 2016). From this camp stem learning approaches such as problem-based learning (e.g., Hmelo-Silver, 1998, Dochy, Segers, van den Bossche, & Gijbels, 2003; Schmidt, 1983), guided inquiry or discovery learning (e.g., Alfieri, Brooks, Aldrich, & Tenenbaum, 2011; DeJong, 2006; DeJong & Lazonder, 2014; Moog, 2014), Invention approaches (Schwartz & Martin, 2004; Roll, Holmes, Day, & Bonn, 2012; Schwartz & Bransford, 1998; Schwartz, Chase, Oppezzo, & Chin, 2011; Belenky & Nokes-Malach, 2012; Glogger-Frey, Fleischer, Grüny, Kappich, & Renkl, 2015; Roll, Alevén, & Koedinger, 2011) and Productive Failure (Kapur, 2010, 2011, 2012, 2014b; Kapur & Bielaczyc, 2011, 2012; Loibl & Rummel, 2014a, 2014b).

Of particular relevance for this dissertation are the two latter learning approaches, namely the Invention approaches and the Productive Failure approach (hereafter PF). In a number of studies and across different samples and domains both learning approaches have shown their benefits for students’ conceptual knowledge acquisition and transfer performance (often without comprising for procedural knowledge) (e.g., PF approach: Kapur, 2010, 2011, 2012, 2014b; Kapur & Bielaczyc, 2011, 2012; Loibl & Rummel, 2014a, 2014b; Invention approach: Schwartz et al., 2011; Schwartz & Martin, 2004; Belenky & Nokes-Malach, 2012). Indeed, students learning with PF or Invention by solving problems prior to receive instruction outperform their counterparts receiving first instruction and then solving problems (i.e., Direct Instruction, DI, problem solving after instruction) on con-
ceptual knowledge and transfer performance. Conceptual knowledge is defined as implicit or explicit understanding about underlying principles and structures of a domain (Rittle-Johnson & Alibali, 1999) and it prevents students from applying rules without further reasoning (at least in the field of mathematics education; Skemp, 1976).

Students learning with an Invention approach are usually first asked to collaboratively (i.e., small groups) invent and evaluate solutions for a yet unknown problem and then receive instruction about how to correctly solve the problem (often followed by time to practice the correct procedure) (e.g., Schwartz & Martin, 2004). In the majority of Invention approaches (e.g., Schwartz & Martin, 2004; Roll et al., 2011; Glogger-Frey et al., 2015) so called contrasting cases reflect a key component of the invention activity. Contrasting cases are examples of different data sets for the same problem that vary only in one deep target feature at a time (Chase, Shemwell, & Schwartz, 2010). By collaboratively discussing, comparing and contrasting different cases during invention students are enabled to activate their prior knowledge and to become aware of deep features of the problem at hand (Loibl, Roll, & Rummel, 2016). In consequence, students’ conceptual knowledge acquisition (Roll et al., 2011) and their transfer performance (Belenky & Nokes-Malach, 2012; Schwartz & Martin, 2004) are supported.

Also in PF students first engage in solving a yet unknown problem typically in small groups prior to receive explicit instruction about how to successfully solve the given problem (e.g., Kapur, 2011, 2012; Kapur & Bielaczyc, 2012; Loibl & Rummel, 2014a, 2014b). During collaborative problem solving students activate and differentiate their prior knowledge by generating incomplete solutions and by discussing strengths and weak points of these solutions (Kapur & Bielaczyc, 2012; Loibl et al., 2016). Typically, the PF instruction builds upon these student-generated solutions as the instructor compares and contrasts student-generated solutions with the canonical solution of the given problem during a class discussion (e.g., Loibl & Rummel, 2014a). By drawing students’ attention to the erroneous features of their own solutions and by comparing and contrasting these erroneous features with the features of the canonical solution the instructor enables students to become aware of gaps in their knowledge and to recognize deep features of the target domain. Again in consequence, students’ acquisition of conceptual knowledge (Kapur & Bielaczyc, 2012) and their transfer performance (Kapur, 2012) are supported.
According to Loibl, Roll, and Rummel’s recent review (Loibl et al., 2016) the effectiveness of PF and Invention approaches rests upon three core learning mechanisms (underlying these approaches). These learning mechanisms are students’ activation and differentiation of prior knowledge (cf. learning mechanism 1), students’ awareness of knowledge gaps (cf. learning mechanisms 2) and students’ recognition of deep features of the target domain (cf. learning mechanism 3). More details about the way the learning mechanisms play out will be presented in chapter 2.1.

While the effectiveness of the PF and Invention approaches is well investigated for middle school, high school and university students, studies with elementary school students using similar approaches with delayed instruction after problem solving show mixed results with regard to students’ conceptual knowledge acquisition when compared with the typical DI control condition (i.e., instruction followed by problem solving) (DeCaro & Rittle-Johnson, 2012; Fyfe, DeCaro, & Rittle-Johnson, 2014; Loehr, Fyfe, & Rittle-Johnson, 2014; Matlen & Klahr, 2013). Despite the commonalities across these studies and Invention and PF studies there are two core differences that might explain why approaches with delayed instruction so far did not consistently show their benefits for young students’ conceptual knowledge: First, the studies conducted with elementary school students did not provide learners with a comparing and contrasting activity either during problem-solving in form of contrasting cases (cf. Invention approach) or in form of an instructor-led comparing and contrasting activity during instruction (cf. PF approach). Second, in contrast to Invention or PF studies conducted with older students, similar studies with elementary school students did not involve peer collaboration during the initial unsupported problem solving phase. The absence of these two design components might explain the mixed results found so far concerning the effectiveness of delaying instruction after problem solving for young students. Indeed, both design components are expected to trigger the aforementioned learning mechanisms underlying the PF and Invention approaches.

For example, when the instructor compares and contrasts student-generated solutions with the canonical solution during the instruction phase of PF he or she enables students to become aware of gaps in their knowledge shown in the often erroneous student-generated solutions (cf. learning mechanisms 2) and to recognize deep features of the target domain by explaining how the canonical solution overcomes the erroneous features (cf. learning mechanism 3). This comparing and contrasting activity might be particularly beneficial for young students because they have lower cognitive capacities (Gathercole,
Pickering, Ambridge, & Wearing, 2004) as compared to older students and might thus have difficulties to focus on the relevant features. In addition, attempting to solve a problem collaboratively helps students to activate and differentiate their prior knowledge by discussing and mutually explaining different solution ideas. As students usually fail to generate the canonical solution for a challenging problem, they need to deal with a potentially frustrating learning experience. In comparison to older students, young students might become more easily frustrated as they probably have lower self-regulation strategies to cope with the specific challenges of problem solving prior to instruction. For example, Landmann and colleagues expect students to be able to learn self-regulatedly at the age of 15-16 years (Landmann, Perels, Otto, Schnick-Vollmer, & Schmitz, 2015). Therefore, in sense of the proverb “a sorrow shared is a sorrow halved”, students’ collaboration with a peer during problem solving might be particularly helpful for young students around the elementary school age.

Against this background the overarching goal of my dissertation was to further test the effectiveness of delaying instruction after problem solving for young students with an improved study design. The improved study design was implemented as a PF approach and included an instruction with an instructor-led comparing and contrasting activity as in light of Loibl and Rummel’s findings (Loibl & Rummel, 2014a) this kind of activity led to higher conceptual knowledge when compared to instruction without such an activity. The improved study design further allowed young learners to collaborate with a peer during problem solving. In contrast to the comparing and contrasting activity during instruction the effects of students’ collaboration on students’ conceptual knowledge have not yet been sufficiently touched in the PF literature. Therefore, I aimed to test for the impact of students’ collaboration on students’ conceptual knowledge by systematically varying the social form of learning during students’ initial problem solving (collaborative learning during problem solving vs. individual learning during problem solving).

In light of Loibl and colleagues recent review the beneficial effect of PF on students’ conceptual knowledge is related to the evolvement of the three aforementioned learning mechanisms leading up to the changes in students’ conceptual knowledge. To test for the effectiveness of delaying instruction after problem solving for young students’ conceptual knowledge thus also required to systematically analyze to which extent these learning mechanisms came to play. A further goal of this thesis is thus to systematically analyze the extent to which all three learning mechanisms evolved once students are enabled to
collaborate during problem solving and to receive instruction with a comparing and contrasting activity.

In order to reach the aforementioned goals I conducted two quasi-experimental studies with students around the elementary school age (i.e., German fourth graders and fifths graders who are typically between 8-12 years old). In both studies I varied the factors timing of instruction with instructor-led comparing and contrasting activity (i.e., problem solving prior to instruction vs. problem solving after instruction) and the social form of learning (collaborative learning during problem solving vs. individual learning during problem solving). The first study served as a feasibility study and was thus set up at a small scale and not with a complete 2x2 design. The three main goals of this study were 1) to develop learning and test material for a younger target group, 2) to develop methods of analysis for detecting underlying learning mechanisms in Study 2, 3) to preliminarily test for the main hypotheses. The Study 2 built upon the findings of Study 1 and was conducted at a larger scale. In Study 2 the two aforementioned factors were systematically varied leading to a complete 2x2 design. In consequence, Study 2 allowed to systematically test for all hypotheses and to analyze the aforementioned learning mechanisms.

In the next chapter I will provide a more detailed overview of the literature concerning the effectiveness of the PF approach and its core learning mechanisms (chapter 2.1). In a second step I will refer to PF-similar studies with young students (chapter 2.2). As these studies differ with regard to two core design components, chapters 2.3 and 2.4 unfold why both components potentially play such an important role for the effectiveness of PF. Chapter 2.5 highlights the need to analyze underlying learning mechanisms and proposes an approach how to do so. The theoretical chapters conclude with a summary of hypotheses (chapter 2.6). In chapter 3 the initial feasibility study (i.e., Study 1) is presented. While chapter 3.1 focuses on the applied methods, chapter 3.2 combines the results and discussion concerning the three aforementioned study goals. In chapter 4 the larger scale Study 2 is presented. Chapter 4.1 illustrates the methods partly being adapted from Study 1 and focusing on analyzing underlying learning mechanisms (see chapter 4.1.4). The results regarding all hypotheses and underlying learning mechanisms are then presented in chapter 4.2 and are discussed in chapter 4.3. Chapter 5 focuses on the general discussion across both studies. While in chapter 5.1 the main results are summarized and discussed, chapter 5.2 points towards limitations and provides an outlook. In chapter 5.3 I will derive pedagogical implications and will conclude the general discussion with a conclusion (see chapter 5.4).
2 Theoretical background

2.1 Productive Failure

As mentioned in the introduction PF is a learning approach in which students first try to solve a yet unknown problem and then receive instruction. In other words, PF usually comprises two learning phases: A problem-solving phase and an instruction phase. For more details about the design of PF please review Kapur and Bielaczyc work to which I mainly refer in this chapter (Kapur & Bielaczyc, 2012): During the initial problem-solving phase of PF students try to collaboratively (i.e., in small groups of two-four students) solve a yet unknown and complex problem. The complexity of a problem results from an interaction between characteristics of the problem and the learner (see also Kapur & Bielaczyc, 2012). With regard to the characteristics of the learners, the main focus lies on students’ prior knowledge. In PF students have not yet been formally introduced into the concepts of the target domain but have already developed an initial understanding about pre-concepts and intuitive ideas of the domain at hand (cf. Kapur, 2012; Loibl & Rummel, 2014a). This means, students do not have sufficient prior knowledge to successfully solve the problem at hand and thus fail by generating incomplete and erroneous solution ideas. With regard to the characteristics of the problem, in PF the given problem allows students to generate and evaluate multiple solutions ideas. In other words, the problem at hand is formulated in such a way that it allows to generate more than a single solution idea. However, in PF, students know that they are not expected to solve the problem successfully (Kapur & Bielaczyc, 2012). This way, the students do not become too easily frustrated with this kind of challenging problem. In addition, they are usually provided with prompts (e.g., you are doing fine. Keep going) designed to maintain students’ motivation. In the subsequent instruction phase, the instructor or teacher then builds upon students’ incomplete and erroneous solution ideas by comparing and contrasting different solution ideas with each other and with the canonical solution (Kapur & Bielaczyc, 2012; Loibl & Rummel, 2014a).

Various studies across different domains and samples with different cultural backgrounds or ability profiles (Kapur, 2010, 2011, 2012, 2014b; Kapur & Bielaczyc, 2011, 2012; Loibl & Rummel, 2014a, 2014b) have consistently demonstrated a superiority of PF over DI (i.e., problem solving after instruction) for students’ conceptual knowledge and partly
for their transfer knowledge. As explained above conceptual knowledge allows students to understand the principles and structures of a domain (Rittle-Johnson & Alibali, 1999).

The effectiveness of PF for students’ conceptual knowledge about the target domain has mainly been ascribed to the three aforementioned core learning mechanisms: 1) Students’ prior knowledge activation and differentiation, 2) students’ awareness of knowledge gaps and 3) students’ recognition of deep features of the target domain (Kapur & Bielaczyc, 2012; Loibl et al., 2016). During the initial problem-solving phase of PF, students activate and differentiate their prior knowledge (cf. learning mechanism 1) by generating and discussing various solution ideas usually in small groups. As students usually fail to come up with the canonical solution by reaching impasses (i.e., failed solution idea) they develop a global awareness about something being wrong with their solution attempts during problem solving (cf. Loibl & Rummel, 2014a; learning mechanism 2). As the incomplete solution ideas might not only comprise erroneous but also some correct solution features, students might already discover some deep features of the target domain during the initial problem-solving phase of PF (cf. learning mechanism 3). During the subsequent instruction phase, students are then enabled to become aware of more specific gaps in their knowledge and to recognize the relevant deep features of the target domain than during problem solving, because the instructor draws students’ attention to the erroneous features of their earlier solution ideas and elaborates on the deep features of the canonical solution. According to Loibl and colleagues (2016) the three learning mechanisms are not independent from one another, but are interconnected in a cascade fashion as learning mechanism 1 paves the way for learning mechanism 2 which in turn paves the way for learning mechanism 3. In other words, in order to be able to recognize deep features of the target domain (cf. learning mechanism 3) students need to become aware of what they already know and what they do not yet know (cf. learning mechanism 2). For this, they have to activate, differentiate and elaborate on their prior knowledge (cf. learning mechanism 1).

To the best of my knowledge there is no published study yet that has implemented the aforementioned PF study design and investigated its effectiveness for learning with students around the elementary school age. There are a few studies implementing not PF, but related learning approaches, in which elementary school students tried to solve problems with delayed instruction, that is, the core design component of the PF approach. I will review these studies on a more detailed level and will derive design implications for the planned studies (see next chapter).
2.2 Delaying instruction for young learners

In the PF-similar studies (see Table 2.1) with elementary school students the study design usually included two (and more) experimental conditions in which students first engaged in problem solving and then received instruction or in which students first received instruction and then engaged in problem solving (i.e., control condition). After problem solving followed by instruction or the other way around (i.e., instruction followed by problem solving) students had to fill out a posttest measuring for students’ conceptual and procedural knowledge. There are thus several commonalities between the studies with elementary school students learning from delayed instruction (see Table 2.1) and PF studies with older students: both kinds of studies implemented delayed instruction as a core feature, employed the same kind of control condition, and used the same kind of learning outcome measures.

However, so far, the PF-similar studies have revealed mixed results concerning the effectiveness of delaying instruction after problem solving on young students’ acquisition of conceptual knowledge. So far only De Caro and Rittle-Johnson (DeCaro & Rittle-Johnson, 2012) found that second graders solving math equivalence problems prior to instruction acquired more conceptual knowledge than second graders solving problems after instruction. In contrast, in two experiments conducted by Loehr and colleagues (Loehr et al., 2014) this finding could not be replicated. For instance, in their first experiment, students who first received instruction and then solved math equivalence problems scored significantly higher at the conceptual knowledge posttest than children who first solved math equivalence problems and then received instruction. Loehr and colleagues (2014) explained the missing superiority of problem-solving prior to instruction over problem-solving after instruction with the fact that students in their first experiment had had no chance to apply their newly accomplished knowledge after instruction. Therefore, in their second experiment—again conducted with second graders—they implemented a knowledge application activity (= so called check phase: revisiting and revising the problems students had solved by using another pen) after problem solving and instruction (and after instruction and problem solving). Again there was no significant difference between students solving problems either prior to or after receiving instruction with regard to their conceptual knowledge. Students who first received instruction, descriptively even reached higher conceptual knowledge posttest scores (and significantly higher procedural knowledge posttest scores). Also Fyfe and colleagues (Fyfe et al., 2014) found no superiority of problem solving prior to instruction to problem solving after instruction. Similarly, Matlen
and Klahr (Matlen & Klahr, 2013) did not find differences between students starting with low guidance followed by high guidance or the other way around when students did not learn to solve math equivalence problems as in the aforementioned studies but about the control of variable strategy for scientific experimentation (in grade 3 in US elementary school).

Table 2.1. Overview about PF-similar studies conducted with elementary school students

<table>
<thead>
<tr>
<th>Study</th>
<th>Age group</th>
<th>Domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Matlen &amp; Klahr, 2012</td>
<td>U.S. third graders</td>
<td>Control of variable strategy (scientific experimentation)</td>
</tr>
<tr>
<td></td>
<td>Average age of sample 9,12 years (N= 57)</td>
<td></td>
</tr>
<tr>
<td>Decaro &amp; Rittle-Johnson, 2012</td>
<td>U.S. second to fourth graders</td>
<td>Math equivalence (meaning of the equal sign)</td>
</tr>
<tr>
<td></td>
<td>Average age of sample 8,5 years (N= 159)</td>
<td></td>
</tr>
<tr>
<td>Fyfe, Decaro &amp; Rittle-Johnson, 2014</td>
<td>U.S. second and third graders</td>
<td>Math equivalence (meaning of the equal sign)</td>
</tr>
<tr>
<td></td>
<td>(N= 122)</td>
<td></td>
</tr>
<tr>
<td>Loehr, Fyfe &amp; Rittle-Johnson, 2014</td>
<td>U.S. second graders</td>
<td>Math equivalence (meaning of the equal sign)</td>
</tr>
<tr>
<td>Experiment 1</td>
<td>7-8 years</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(N= 41)</td>
<td></td>
</tr>
<tr>
<td>Loehr, Fyfe &amp; Rittle-Johnson, 2014</td>
<td>U.S. second graders</td>
<td>Math equivalence (meaning of the equal sign)</td>
</tr>
<tr>
<td>Experiment 2</td>
<td>7-8 years</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(N= 47)</td>
<td></td>
</tr>
</tbody>
</table>

Summing up, despite the aforementioned commonalities between the studies with elementary school students learning from delayed instruction, and PF studies with older students there are two core differences between both groups of studies, which might explain why, so far, the young students did not consistently benefit from solving problems prior to receiving instruction: the form of instruction during the instruction phase, and the social form of learning during the problem-solving phase. I will explain both core differences subsequently. In this context, I will highlight the need to adapt the study design of the aforementioned studies to a PF design. This way, I will be able to shed further light on the effectiveness of delaying instruction after problem solving for young students around elementary school age.
2.3 Form of instruction

A key design component of PF is an instructor-led comparing and contrasting activity between the incomplete student-generated solutions and the canonical solution during the instruction phase. In contrast to PF studies conducted with older students, none of the aforementioned studies conducted with younger students included a comparing and contrasting activity during the instruction phase. However, this PF design component is in light of findings by Loibl and Rummel (Loibl & Rummel, 2014a) and their recent review (Loibl et al., 2016) particularly important for the evolvement of two core learning mechanisms underlying PF (i.e., students’ awareness of knowledge gaps and the recognition of deep features of the target domain) and for students’ conceptual knowledge acquisition in turn. The way students’ awareness of knowledge gaps (cf. learning mechanism 2) and students’ recognition of deep feature (cf. learning mechanism 3) are supported by an instructor-led comparing and contrasting activity will be explained step-by-step.

More precisely, in line with the impasse-repair-reflect process (VanLehn, 1999) and Chi’s work (2000) the comparing and contrasting activity helps students to first become aware of their impasses or knowledge gaps (cf. learning mechanism 2) and then helps them to repair their initial suboptimal mental model by contrasting the erroneous solution with the canonically correct solution (for further details see Loibl et al., 2016; Loibl & Rummel, 2014a). In consequence, the awareness of one’s own knowledge gaps lays the foundation for deep processing of the canonical solution (e.g., VanLehn, Siler, & Murray, 2003) as explained by Loibl and colleagues (2016).

The importance of students’ awareness of knowledge gaps for the effectiveness of PF receives support from two other angles: In a quasi-experimental study conducted by Nachtigall, Rummel and Serova (2017) students solved problems either prior to or after instruction in the domain of social and educational science methodologies. They found that students overestimated the quality of their solution ideas. This way, students were unable to evaluate their (incomplete) solutions as failed and thus did not engage in reflecting about their failure thereafter. In addition, to some extent even the findings around PF studies with vicarious failure (Kapur, 2014a, 2014b) showing a superiority of the “classic” PF condition over a vicarious failure condition underlines the importance of students’ awareness of knowledge gaps as the awareness of failure probably is more likely and stronger by generating own erroneous solution ideas rather than by learning from a failed example (i.e., vicarious failure condition).
In addition, when the instructor compares different erroneous student solutions with the canonical solution, explains why these solutions are erroneous and how the canonical solution overcomes these deficits he or she also enables students to recognize the deep features of the target domain. The importance of noticing and recognizing deep features is well established in the expertise literature (Chi, Feltovich, & Glaser, 1981) because it facilitates experts to differentiate relevant from irrelevant (i.e. superficial) features of a given problem. This way, students are enabled to organize the target knowledge around the deep features and to more easily adapt and transfer it to new situations (cf. Loibl et al., 2016).

Beyond the theoretical rationale, Loibl and Rummel (Loibl & Rummel, 2014a) provide empirical evidence for including an instructor-led comparing and contrasting activity within the instruction phase of PF for facilitating both aforementioned learning mechanisms. By clearly distinguishing between the timing of instruction (instruction prior to or after problem solving) and the form of instruction (with or without comparing and contrasting activities) Loibl and Rummel (Loibl & Rummel, 2014a) showed that the beneficial effect of problem solving prior to instruction in comparison to instruction prior to problem solving only come to play when students received instruction with an instructor-led comparing and contrasting activity. In a recent review of the PF literature, Loibl and colleagues (2016) summarize that only those implementations of problem-solving prior to instruction approaches that include a comparing and contrasting activity were highly effective.

Interestingly, literature regarding learning with examples emphasizes the need to compare not just different correct examples or different erroneous examples with each other but both correct and erroneous examples with one another. For example, Durkin and Rittle-Johnson (Durkin & Rittle-Johnson, 2012) showed that students learning with incorrect and correct examples gained more conceptual knowledge as compared to students learning with just one type of examples.

Because the effectiveness of a comparing and contrasting activity during instruction is well investigated and is known to facilitate students’ conceptual knowledge acquisition (Loibl & Rummel, 2014a), the first goal of this thesis is to include this crucial component in the design of my studies. This way and in line with the aforementioned PF-similar studies, I will be able to further investigate the effects of delaying instruction after problem solving for young students (with adapted instruction). More precisely, I hypothesize
problem-solving prior to instruction (i.e., PF) leads to higher conceptual knowledge acquisition by young students, as compared to problem-solving after instruction (i.e., DI) (Hypothesis 1).

2.4 Social form of learning

Again in contrast to the majority of PF studies (e.g., Kapur, 2011, 2012; Kapur & Bielaczyc, 2012; Loibl & Rummel, 2014a, 2014b) the aforementioned studies with elementary students do not only differ with regard to the form of instruction, but also with regard to the social form of learning during the initial problem-solving phase. While in the PF studies with older students, students usually try to solve the yet unknown problem collaboratively (i.e., in small groups), the aforementioned studies with the younger elementary students (see also Table 2.1) typically had them work individually or together with the experimenter. However, considering Kapur’s and Bielaczycc’ work (2012), small group collaboration is another crucial PF design component that specifically facilitates students’ prior knowledge activation and differentiation by enabling them to generate and discuss different solution ideas with their peers during the problem-solving phase (cf. learning mechanisms 1).

It is well investigated that students’ prior knowledge activation is important for students’ learning (e.g., Sweller, van Merrienboer, & Paas, 1998) as the integration of new knowledge into existing schemas (see also Linn, 1995) more likely comes to play when the existing (relevant) schemas are activated (e.g., Sweller et al., 1998) and are thus prepared for being modified (cf. Loibl et al., 2016). As explained in Loibl and colleagues recent review (2016), for example, in two experiments Schmidt and colleagues (Schmidt, Volder, Grave, Moust, & Patel, 1989) showed understanding an explanatory text about natural phenomena is facilitated by preceding small group discussion (which to some extent resembles collaborative problem solving followed by instruction).

Further support for enabling students to collaborate during problem solving comes from the broader literature on learning with failure (e.g., Piaget’s and Brown’s cognitive disequilibrium, 1985; Van Lehn’s impasse-driven learning, 1988). According to Tawfik and colleagues (Tawfik, Rong, & Choi, 2015) learning with failure happens in a cyclic way and covers the following cognitive steps learners need to engage in: developing an initial mental model about the given problem, experiencing failure, analyzing for reasons for failure as the initial mental model is being challenged, generating and implementing new
solutions, and finally developing an adapted mental model about the problem (i.e., key steps of failure-based problem solving). To adapt to these requirements collaboration might play a supportive role. For instance, collaboratively learning students as compared to individually learning students are during the development of an initial mental model (first cognitive step of the failure cycle) enabled to receive feedback not only from one instance (i.e., the problem at hand) but additionally from a second (and third) instance, namely their learning partner(s). Therefore, collaboratively learning students are enabled to make adjustments during the development of the solution idea rather than after failure. Although students’ collaboration probably does not prevent them from experiencing failure, it possibly helps them to evaluate their failure and to deal with analyzing for failure by mutually explaining and discussing why a solution idea failed (e.g., as the solution idea at hand probably misses a deep feature of the target domain). In consequence, through students’ collaboration elaborative processes (e.g., Teasley, 1995) and sense-making activities (e.g., Koedinger, Corbett, & Perfetti, 2012; Asterhan & Schwarz, 2009) are triggered which are known to support students’ conceptual knowledge acquisition.

Further support for the supportive role of collaborative learning during students’ initial problem solving comes from broader literature on collaborative learning. Indeed, also in this literature (e.g., Sears & Reagin, 2013) the “good fit” between students’ collaboration as one social form of learning and characteristics of the problem at hand are emphasized. More precisely, in a study conducted by Sears and Reagin (2013) students either individually or collaboratively tried to solve a complex task. Similar as the PF literature also Sears and Reagin (2013) understand complexity as an interaction between features of the problem and features of the student such as prior knowledge. While students with higher prior knowledge (i.e., accelerated class) benefitted from learning individually to solve less complex tasks, students with lower prior knowledge benefitted from learning collaboratively on a task that was complex for them. In addition, also in a study conducted by Mullins and colleagues (Mullins, Rummel, & Spada, 2011) students’ collaboration during working on more conceptually oriented learning material (to some extent assembling solving a complex problem in a PF setting) particularly facilitated students’ conceptual knowledge acquisition. Therefore, students’ collaborative learning during the initial problem-solving phase of PF might be particularly beneficial for students’ conceptual knowledge acquisition.

Nevertheless, research on collaborative learning has also elicited that the benefits of collaborative learning might in light of the interaction paradigm (Dillenbourg, Baker, Blaye,
& O’Malley, 1996) only come to play when students engage in fruitful collaborative processes. To ensure such fruitful collaborative processes students thus need to be supported as further research on collaborative learning revealed (e.g., Fischer et al., 2013). One way to support students’ fruitful collaborative processes is to provide them with a group goal serving as motivational incentive and a role script to help students to actually engage in team work and mutual support (cf. King, 2007; Slavin, 1995).

Against this background my second goal is to examine whether also young students benefit from collaborating with a peer during the problem-solving phase of PF. As so far little is known about the effects of older and younger students’ collaboration I systematically vary the social form of learning during problem solving. More precisely, I aim to test whether students learning collaboratively (i.e., in a dyad) during the initial problem-solving phase of PF (followed by instruction) acquire more conceptual knowledge as compared to students learning individually (Hypothesis 2). To facilitate fruitful collaborative processes between learning partners (rather than to facilitate fruitful cognitive problem solving processes) (Westermann & Rummel, 2012) collaboratively learning students will be supported.

### 2.5 Analyses of learning mechanisms underlying Productive Failure

To test for the effectiveness of delaying instruction after problem solving in form of PF (i.e., including an instruction with a comparing and contrasting activity and students’ collaboration during problem solving) requires in light of Loibl and colleagues recent review (2016) not just to examine students’ conceptual knowledge as learning outcome but to analyze the three aforementioned learning mechanisms which are expected to lead to the beneficial effect on students’ conceptual knowledge. In consequence, one needs to systematically analyze to which extent these learning mechanisms indeed evolved during the preceding learning process (i.e., during the problem-solving and instruction phase). Therefore, an additional goal of this thesis is to systematically investigate how these core learning mechanisms come to play once students are enabled to collaborate during problem solving and to receive instruction with a comparing and contrasting activity. However, despite the relevance of systematically analyzing all three learning mechanisms when investigating the effects of delaying instruction after problem solving on students’
conceptual knowledge, so far little is known about how to evaluate to which extent students activated and differentiated their prior knowledge (cf. learning mechanism 1), became aware of gaps in their knowledge (cf. learning mechanism 2) and recognized deep features of the target domain (cf. learning mechanism 3) and how to evaluate to which extent all three learning mechanisms interact with one another (i.e., cascade way) (Loibl et al., 2016). My work thus displays a first step towards a systematic analysis of underlying learning mechanisms.

By reviewing the literature about PF and related approaches I aim to identify potential indicators for the evolvement of the learning mechanisms. Some of these indicators have previously been used not for systematically analyzing all three learning mechanisms but for analyzing single learning mechanisms or single aspects of these mechanisms. In this context, I will further explain how these potential indicators might be affected by the improved study design (i.e., instruction with comparing and contrasting activity and students’ peer collaboration).

### 2.5.1 Students’ prior knowledge activation and differentiation

In previous PF studies (Kapur, 2012; Kapur & Bielaczyc, 2012) typically the number of solution ideas students generated during the initial problem-solving phase of PF was used as an indicator for the extent to which students activated and differentiated their prior knowledge (Kapur uses the term representation and solution methods, RMS, to denote students solution ideas). Indeed, some studies showed the more solution ideas students generated or likewise the higher the variety of different representation and solution methods was the higher was their learning outcome (Kapur, 2012, 2014b; Kapur & Bielaczyc, 2012).

In addition, as students’ collaboration during the initial problem-solving phase is meant as a central PF component for triggering learning mechanism 1, it needs to be investigated whether the number of student-generated solution ideas differs between collaboratively and individually learning students (in PF). More precisely, as students in a collaborative setting are forced to verbalize and explain their solution ideas and might thus stimulate each other’s solution process they probably come up with a higher number of solution ideas as compared to students engaging in individual problem solving. Therefore, Hypothesis 3 is students learning collaboratively during the problem-solving phase of PF generate more solution ideas as compared to individually learning students (in PF).
As in light of the interaction paradigm (Dillenbourg et al., 1996) the actual collaborative processes are more important for students’ learning than the social form of learning per se, I aim to shed more light on the interplay between different kinds of collaborative processes and students’ prior knowledge activation and differentiation (i.e., number of student-generated solutions) and students’ conceptual knowledge acquisition. While, for example, students’ collaborative off-task processes and coordinative processes such as managing the communication processes probably hampers the generation of solution ideas (cf. learning mechanism 1) and students’ conceptual knowledge acquisition in turn, interactively building upon the learning partner’s previous solution idea (i.e., number of solution ideas) probably facilitates a broad prior knowledge activation (cf. learning mechanism 1) and students’ acquisition of conceptual knowledge in turn.

Further support for the interplay between different kinds of learning processes and students’ learning outcome (as here students’ conceptual knowledge) can be found in the Interactive-Constructive-Active-Passive framework (ICAP-framework, Chi, 2009; Chi & Wylie, 2014). The ICAP-framework aligns different kind of overt learning processes (not just collaborative learning processes) in a hierarchical order to underlying cognitive processes which in turn are differently beneficial for students’ learning. For instance, in line with the ICAP-framework interactive processes by interactively building upon each other’s contributions trigger so called joint creating processes which in turn facilitate students’ learning. In addition, such interactive processes are more beneficial for students’ learning as compared to constructive processes (e.g., giving a self-explanation) triggering “only” creating processes. Both interactive and constructive processes are more beneficial for students learning as compared to active processes (e.g., reading a text) which “only” trigger attending processes and thus do not facilitate deep learning. All three types of learning processes (i.e., interactive, constructive and active) are more beneficial for learning as passive processes that only imply the physical presence of a learner (without further learner engagement).

2.5.2 Students’ awareness of knowledge gaps

By implementing a comparing and contrasting activity during the instruction phase the (study) design specifically addresses (young) students’ awareness of knowledge gaps (and recognition of deep features of the target domain, i.e., learning mechanism 3). However, the benefits of the comparing and contrasting activity with regard to learning mecha-
anism 2 might only develop their full potential when students actually generated *erroneous and incomplete* solutions by themselves and thus had the chance to become aware of their failure already during the initial problem-solving phase of PF. Therefore, the quality of student-generated solutions serves as a proxy for students to become aware of gaps in their knowledge. In addition and in line with for example Wiedmann and colleagues’ work (2012) and Loibl and Rummel’s work (Loibl & Rummel, 2014b) it needs to be investigated how the solution quality is linked to students’ conceptual knowledge acquisition.

Against the background on research on collaborative learning students’ collaboration during the initial problem-solving phase might not only trigger learning mechanism 1 but additionally learning mechanism 2 (and partly 3): As collaboratively learning students are enabled to discuss strengths and weak points of a solution and to receive mutual feedback, they are enabled to improve a solution by discussion. Therefore, I hypothesize collaboratively learning students to generate higher quality solutions when compared to individually learning students (i.e., Hypothesis 4).

In addition and again in line with the interaction paradigm (Dillenbourg et al., 1996) a further goal is to analyze how different kinds of collaborative processes are linked to the solution quality. For example, by interactively building upon each others’ contributions and by constructively explaining solutions to each other it might be that collaboratively learning students already resolve some of their knowledge gaps during the problem-solving phase. Following this line, I aim to investigate how different kinds of collaborative processes are connected to the solution quality.

### 2.5.3 Students’ recognition of deep features of the target domain

By implementing a comparing and contrasting activity during the instruction phase I intend to address both students’ awareness of knowledge gaps (i.e., learning mechanism 2) and additionally students’ recognition of deep features of the target domain (i.e., learning mechanism 3). From research on expertise (Chi et al., 1981) we know domain experts in contrast to novices are able to recognize the deep and important features of the target domain and are thus more capable to differentiate between irrelevant superficial features and the relevant deep features. Generally, through the comparing and contrasting activity novices are able to approach the problem more like an expert as the instructor specifically draws students’ attention to the deep rather than the superficial features of the target domain. However, students who engaged in problem solving prior to instruction are less
novices by the timing of this specific instruction as compared to students who first re-
ceived instruction and then solved problems (Kapur, 2016). In consequence, the per se 
beneficial form of instruction might be less beneficial for students receiving instruction 
first and then engaging in problem solving. For this, students learning with PF as com-
pared to students learning with DI might be more able to recognize deep features of the 
target domain that are required to successfully answer the conceptual knowledge posttest 
as the conceptual knowledge posttest specifically asks for these deep features) (cf. Hy-
pothesis 1).

2.6 Summary of hypotheses

The overarching goal of this dissertation is to further test the effectiveness of delaying 
instruction after problem solving for young students with an improved study design. As 
compared to previous PF-like studies, two major changes to the experimental design were 
implemented: An instruction with an instructor-led comparing and contrasting activity 
and students’ collaboration during problem solving. The two main hypotheses are:

Hypothesis 1: Problem solving prior to instruction with instructor-led comparing and 
contrasting activity (i.e., PF) leads to higher conceptual knowledge acquisition in young 
students, as compared to problem solving after instruction (i.e., DI) with an instructor-led 
comparing and contrasting activity. To test for Hypotheses 1, I compare students solving 
problems prior to receiving instruction (i.e. PF) with students solving problems after re-
ceiving instruction (i.e., DI).

Hypothesis 2: Students learning in a dyad during the initial problem-solving phase of PF 
(followed by instruction) acquire more conceptual knowledge as compared to students 
learning individually. To test for Hypothesis 2, I compare students learning collabora-
tively (i.e., in a dyad) during the initial problem-solving phase of PF followed by instruc-
tion with students learning individually during problem-solving followed by instruction.

While these two hypotheses refer to students’ conceptual knowledge as learning outcome, 
the following hypotheses refer to the learning mechanisms and processes leading up to 
this outcome.

To shed light on learning mechanism 1 (i.e., students’ prior knowledge activation and 
differentiation) I investigate the number of student-generated solution ideas and how this 
number is linked to students’ conceptual knowledge acquisition. In this context, it will be 
examined whether students learning in a dyad during the initial problem-solving phase of
PF generate more solution ideas as compared to students learning individually (cf. Hypothesis 3). To test Hypothesis 3, I compare whether students learning collaboratively or individually generate more solution ideas. In context of learning mechanism 1, I additionally focus on the interplay between different kinds of collaborative processes and the number of student-generated solution ideas and the interplay between collaborative processes and students’ conceptual knowledge.

To investigate learning mechanism 2 (i.e., students’ awareness of knowledge gaps) I analyze the solution quality that serves as proxy for the extent of students’ knowledge gaps. In this context, I examine how the solution quality is linked to students’ conceptual knowledge. It will further be investigated whether students learning in a dyad during the initial problem-solving phase of PF generate higher quality solutions as compared to students learning individually (cf. Hypothesis 4). To investigate Hypothesis 4, I test whether collaboratively learning students generate qualitatively better solutions as compared to individually learning students. In addition, I examine how different kinds of collaborative processes are linked to the solution quality.

In order to investigate the extent to which the learning mechanism 3 (i.e., students’ recognition of deep target features) evolves I rely on the conceptual knowledge posttest as it specifically requires students to recognize and understand deep features of the target domain.

Summing up, to investigate these hypotheses and evaluate the aforementioned learning mechanisms underlying PF two quasi-experimental field studies were conducted. Both studies varied the factors social form of learning (collaborative learning during problem solving vs. individual learning during problem solving) and timing of instruction with instructor-led comparing and contrasting activity (problem solving prior to instruction vs. problem solving after instruction).

As to the best of my knowledge there are no published PF studies with students around elementary school age yet and therefore no established learning and test materials for this young age group, Study 1 served as a feasibility study and had three goals: The first goal of Study 1 was to develop learning and test material that is appropriate for the younger age group. In parallel to the development of learning and test material the second goal of Study 1 was to design methods of analysis for investigating the extent to which students activate and differentiate their prior knowledge (cf. learning mechanism 1), become aware of gaps in their knowledge (cf. learning mechanism 2) and recognize deep features
of the target domain (cf. learning mechanism 3). To do so, the methods of analysis will rely on the number of student-generated solutions, the solution quality and students collaborative processes. These methods of analysis were then applied to the respective data collected in Study 2. The third and last goal of Study 1 was to preliminary test for the two main hypotheses (i.e., Hypotheses 1 and 2). To reach these goals Study 1 was implemented with three experimental conditions (condition 1: collaborative problem solving prior to instruction, PF; condition 2: individual problem solving prior to instruction, PF; condition 3: collaborative problem solving after instruction, DI) and was conducted at a small scale.

Building upon the findings of Study 1, Study 2 then systematically tested for all aforementioned hypotheses and examined to which extent the aforementioned learning mechanisms underlying PF came to play. For this, Study 2 was conducted with a much larger sample size and a complete 2x2 design that more systematically varied the factors timing of instruction with comparing and contrasting activity (problem solving prior to instruction, PF vs. problem solving after instruction, DI) and social form of learning (collaborative learning during problem solving vs. individual learning during problem solving).
3 Study 1

As explained above the goals of this study were threefold: 1) to develop the learning and test material, 2) to develop methods of analysis for detecting underlying learning mechanisms and 3) to preliminary test for the two main hypotheses regarding the effects of the timing of the re-designed instruction (i.e., with an instructor-led comparing and contrasting activity) and the effect of students’ collaborative learning during the initial problem-solving phase of PF. To reach these goals a quasi-experimental study was conducted. As recruiting a big study sample often comes with high costs this feasibility study was conducted at a small scale and thus not with a complete 2x2 experimental set up.

3.1 Method

In the following sections, the methods developed for the young target group will be described. Particular attention will be payed to outlining the central design components of the learning materials.

3.1.1 Design

A quasi-experimental study with three experimental conditions was implemented. Classes were randomly assigned to conditions. In order to shed further light on the effectiveness of delaying instruction after problem solving, two conditions were implemented in which students first tried to solve a problem and then received instruction (i.e., PF conditions), and a third condition was implemented in which students first received instruction and then collaboratively solved a problem (i.e. DI-Coll condition) (cf. hypothesis 1). The instruction included the PF-typical comparing and contrasting activity. In order to test for Hypothesis 2, the social form of learning during students’ initial problem-solving was varied between the two PF conditions: Students either tried to collaboratively (i.e., PF-Coll condition) or individually (i.e., PF-Ind condition) solve a problem. All three experimental conditions included two learning phases. Table 3.1 provides an overview of the three experimental conditions and the sequence of the respective learning phases per condition.
Table 3.1. Overview of experimental conditions and learning phases in Study 1

<table>
<thead>
<tr>
<th></th>
<th>Learning phase 1</th>
<th>Learning phase 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>PF-Coll condition</td>
<td>Collaborative problem solving</td>
<td>Instruction</td>
</tr>
<tr>
<td>PF-Ind condition</td>
<td>Individual problem solving</td>
<td>Instruction</td>
</tr>
<tr>
<td>DI-Coll condition</td>
<td>Instruction</td>
<td>Collaborative problem solving</td>
</tr>
</tbody>
</table>

3.1.2 Sample

A total of 52 German fourth grade elementary school students participated in this study, with an average age of nearly 10 years ($M=9.98; SD=.68; N=51$). The recruited students were approximately equally old as the students in previous studies that implemented problem solving prior to instruction approaches (Loehr et al., 2014: US second graders who are typically between 7-8 years old both experiments; Fyfe et al., 2014: US second and third graders with an average age of 8.2 years; DeCaro & Rittle-Johnson, 2012: US second to fourth graders with an average age of 8.5 years; Matlen & Klahr, 2013: US third graders with an average age of 9.12 years). One child did not indicate his age. The calculation of the average age was therefore based on 51 rather than 52 students. In line with the German standards for the core curriculum (Ministerium für Schule und Weiterbildung des Landes Nordrhein-Westfalen, 2008) and previous PF studies (cf. Kapur, 2012; Loibl & Rummel, 2014a), students had not yet acquired any formal knowledge about the target domain at the time of the study (but had developed pre-concepts of the target domain). Across all experimental conditions parental consent for students’ involvement was obtained for all participating students.

3.1.3 Learning material

Target domain and conceptual knowledge features

Thematically, Study 1 (as well as Study 2) was embedded in the topic of fractions. Learning fractions is very important as fractions performance at the age of 10 predicts future math performance halve a decade later (Siegler et al., 2012). During both learning phases, the aim was for students to acquire conceptual knowledge about equivalence of fractions.
Students’ conceptual knowledge about equivalence of fractions comprised different target features (or likewise concepts), which to some extent are interconnected. The first feature required students to understand the relation between the size and the number of fractional parts within a given fraction (cf. relation-feature). The second feature built upon the first one as it required students to equally and fairly divide fractions amongst persons or groups of persons (cf. division-feature). Therefore, students needed to understand for example that a person who receives 2/4 of pizza has the same amount of pizza as a person who receives 4/8, although 4/8 is a fraction with a greater numerator and denominator (cf. common preconceptions of students at the transition between grade 5 and 6; Padberg, 2002). The third feature focused on the comparison between two given fractions for identifying the presence or absence of fraction equivalence (cf. comparison-feature). In order to be able to compare two fractions (i.e., with different denominators and numerators) with each other, students needed to understand the necessity to first find a common denominator. Only this way students were able to “mathematically argue” for the existence or absence of fractions equivalence rather than “estimating or seeing” it by for example putting fractional parts together and estimating whether they were equivalent or not. According to the German standards of the core curriculum (Ministerium für Schule und Weiterbildung des Landes Nordrhein-Westfalen, 2008), being able to mathematically argue with mathematical concepts is a key competence students should have acquired at the latest by the end of grade 4.

Design elements of the fractions problem and instruction

Across all three experimental conditions, the design of the equivalent fraction problem built upon the three conceptual knowledge features about equivalent fractions. It is a story problem about dividing and comparing pizzas for a group of boys and of girls at a child’s birthday party. As illustrated in Figure 3.1, the story problem required students to undertake three problem-solving steps which were not marked as such for the students. In the first and second problem-solving step, students are asked to identify the number and the size of pizza pieces each boy or girl receives (i.e., amount of pizza). Therefore, they needed to divide equally and fairly by giving each boy and girl the same number of (equally sized) pieces (cf. deep features 1 and 2). In the third problem-solving step, students had to compare the two amounts of pizza in order to identify the presence or absence of fraction equivalence. As the numerator and denominator of the pizza for the boys and the numerator and denominator of the pizza for the girls usually differed, students had to
understand the need to first find a common denominator to mathematically identify and argue for equivalence by either expanding or reducing one of the two given fractions (cf. deep feature 3). The problem illustrated in Figure 3.1 served as a template for the design of the problem in the PF-conditions and the DI-condition.

Note the design of the equivalent fraction problem built upon an equivalent fraction task designed by a German mathematics education expert for learning and teaching fractions, namely Prof. Friedhelm Padberg (see appendix 7.1.1). By using an equivalent problem that so closely resembled the task created by Padberg (Padberg, 2002, p. 57), I aimed to not only consider PF design requirements but additionally mathematics education literature to increase the likelihood of the problem’s understandability for the young students.

![Figure 3.1. Basic template of the equivalent fraction problem across experimental conditions](image)

In line with the equivalent fraction problem, the instruction also built upon the three aforementioned deep features of equivalent fractions (i.e., relation-, division- and comparison-feature). The instruction was designed as a class discussion. The instructor, compared and contrasted different student solutions to the canonical solution for the problem shown in Figure 3.2, in which all three deep features were used. In line with previous PF studies, the instruction built upon typical erroneous and incomplete students solutions. By using typical incomplete student solutions which were previously collected in a pilot study (Study 0, N=25) instead of the actual answers given by students in Study 1, the instruction was kept constant across all three experimental conditions (cf. Loibl & Rummel, 2014a).
Apart from the instructor-led comparing and contrasting activity, the design of the instruction further takes into account how fractions are taught and learnt on a more general level. The Rational Number Project (e.g., Cramer, Post, & delMas, 2002), a research project that has been investigating learning and teaching fractions for over two decades, has shown that including estimation questions prior to using formal fractions problem facilitates learning. Therefore, a further design component of the instruction was to ask students to estimate whether girls or boys would receive more or less than an entire pizza at the beginning of the instruction (cf. 3rd problem-solving step). In line with mathematics education literature and literature about learning fractions (e.g., Cramer et al., 2002; Rau, Aleven, & Rummel, 2009; Ainsworth, Bibby, & Wood, 2002) a further design element of the instruction was to include graphical representations of fractions (i.e. circle representation). Including graphical representations was also based on the results of the aforementioned pilot study (Study 0; N= 25) in which most student solutions included graphical representations. By building upon the solutions with representations during the instruction phase, the aim was to connect to students’ (informal) prior knowledge and their so called *lebensweltliche experience* (Husserl & Biemel, 1976).

In line with Loehr and colleagues (2014), the instruction phase also included opportunities for students to apply their newly accomplished knowledge about the aforementioned deep target features. Students could practice to equally and fairly divide a given amount of pizza amongst a group of children (cf. relation- and division- feature) and practice to compare two given fractions to each other (cf. comparison-feature).

For the design of the equivalent fraction problem and the instruction, I collaborated with mathematics education experts and teachers, pilot tested and iteratively improved different versions of the equivalent fraction problem and the instruction. This way, I aimed to ensure a mapping between the research goals on the one hand and characteristics of the young target group on the other hand. Detailed descriptions of the actual learning material for each condition will be described in the next chapters.

**Problem-solving phase within Productive Failure and Direct Instruction**

Across all experimental conditions, students engaged in solving a problem that built upon the problem template shown in Figure 3.1 (and they also received the same instruction). The only difference between the problems students were asked to solve in the PF-conditions and the DI-condition lies in the actual numbers being used within the problems.
In both PF-conditions, students tried to solve the problem shown in Figure 3.2. In three problem-solving steps, they had to identify the amount of pizzas for each boy and girl respectively (cf. problem-solving step 1 and 2) and to compare both amounts of pizza (cf. problem-solving step 3). As the young students had not yet had acquired any formal knowledge about the underlying target features, the equivalent fraction problem was rather complex for them. In line with the PF design requirements described in the introduction, the problem enabled students to find multiple solutions ideas by calculating a solution, making use of a graphical representation (i.e. circle with different denominators and numerators), drawing various solution ideas, identifying equivalents of given fractions or by logical reasoning.

To prevent frustration, at the beginning of the study the instructor made clear to the PF students it was not expected they would successfully solve the problem. Additionally, to maintain students’ motivation, the instructor provided them with motivational prompts (e.g., you are doing fine, keep going) during the problem-solving process. To obtain insights into the collaborative problem-solving process, students were videotaped during collaborative problem-solving.

In line with the problem-solving phase of the PF-conditions, students from the DI-Coll condition (preceded by instruction) also tried to solve a story problem that built upon the template shown in Figure 3.1. Students from the DI condition also had to compare two amounts of pizzas for a group of boys and a group of girls at a child’s birthday party. To
solve the problem illustrated in Figure 3.3, the DI-students also needed to engage in three problem-solving steps (i.e., isomorphic problem for the DI-condition). As the problem-solving phase in DI only took place after the instruction, students were not expected to generate multiple solutions ideas.

Marc ordered pizzas for his birthday dinner: For the four boys he ordered two “pizza Salami” and for the eight girls he ordered four “pizza Hawaii”. The boys share the four “pizza Salami” equally among each other. Also the girls share the two “pizza Hawaii” equally among each other.

Who gets more: Does a boy or a girl receive a greater amount of pizza?

Figure 3.3. The equivalent fraction problem within the problem-solving phase of Direct Instruction.

Instruction phase within Productive Failure and Direct Instruction

The instruction was constant across all three experimental conditions. As a starting point of the instruction phase, the instructor asked students the aforementioned estimation question (i.e., whether boys or girls receive a greater, smaller or equal amount of pizza). As illustrated in Table 3.2, the instructor then compared and contrasted three typical erroneous student solutions to the canonical solution for problem-solving step 1 (i.e., dividing the four pizza Salami amongst the six boys).

With regard to student solution 1, the instructor drew students’ attention to the fact that the pizza for the boys is neither equally nor fairly divided as none of the pizza pieces goes through the central point and thus differs in size (cf. relation- and division-feature). This way, students should become aware of their failure (i.e., the solution did not include drawing pizza points through the central point) and students should simultaneously understand why passing through the central point is a prerequisite to generate the canonical solution.

Although in student solution 2 (dividing into quarters) and 3 (dividing into eights) the pizzas were divided by going through the central point, they are still not equally and fairly divided amongst the boys as some boys receive a higher or lower amount of pizza. Students should become aware of the limitations of the illustrated student solutions as there
is always a remainder when dividing pizza pieces (solution 2: 16 divided by 6 = 4, remainder is 4; solution 3: 32 divided by 6 = 5, remainder is 2). Then, the instructor drew students’ attention to the relation between the number of boys (i.e., six boys) and the chosen division into quarters (in solution 2) and eights (in solution 3). In solution 2 the chosen division is lower than the number of boys, while in solution 3 the chosen division is higher than the number of boys. Against this background, the instructor presents the canonical solution for this problem-solving step by dividing the four pizzas into sixths, counting the total number of pieces and dividing the total number (i.e., 24) by the number of boys (i.e. 24 / 6 = 4). As 24 divided by 6 leads to 4 pieces for each boy, the pizza is equally and fairly divided amongst all boys. Finally, in order to demonstrate and document the result for the first problem-solving step, the instructor drew a circle representation of a pizza, divided it into sixths and marked four pieces out of the total of six pieces.

After the instructor compared and contrasted the solutions generated for problem solving step 1, students had the opportunity to apply their newly accomplished knowledge about this deep target feature (i.e., division-feature). Students were asked to again solve problem-solving step 2 that is isomorphic to problem solving step 1. Students were provided with the worksheet illustrated in Figure 3.4 and were instructed to consider the relation between the number of girls by dividing the two pizzas (for the three girls). For problem-solving step 2, then the instructor again demonstrated and documented the results by drawing a pizza representation, dividing it into thirds and marking two pieces out of three pieces to show the number of pieces girls receive.
Table 3.2. Comparing and contrasting activity in the instruction phase for problem-solving step 1

<table>
<thead>
<tr>
<th>Examples for typical erroneous student solutions and the canonical solution</th>
<th>Activity of the instructor in the role of the teacher</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Student solution 1:</strong></td>
<td><strong>Step 0:</strong> Estimation question: Does a boy (and girl respectively) get more or less than an entire pizza?</td>
</tr>
<tr>
<td></td>
<td><strong>Step 1:</strong> Drawing student solution 1, counting the number of pieces and highlighting the absence of drawing through the central point</td>
</tr>
<tr>
<td></td>
<td><strong>Step 2:</strong> Drawing student solution 2 and counting the number of pieces of pizza (=16)</td>
</tr>
<tr>
<td></td>
<td><strong>Step 3:</strong> Dividing 16 pieces of pizza by 6 boys which results in 2 pieces for each boy with a remainder of 2 pieces</td>
</tr>
<tr>
<td></td>
<td><strong>Step 4:</strong> Guiding students’ attention to the goal of dividing equally and fairly which is not yet reached by cutting the pizzas into quarters</td>
</tr>
<tr>
<td></td>
<td><strong>Step 5:</strong> Finding another way of cutting the pizza into pieces</td>
</tr>
<tr>
<td><strong>Student solution 2:</strong></td>
<td><strong>Step 6:</strong> Repeating Steps 2-5 for student solution 2</td>
</tr>
<tr>
<td></td>
<td><strong>Step 7:</strong> Asking how the number of boys is linked to the number of pieces of pizza?</td>
</tr>
<tr>
<td><strong>Student solution 3:</strong></td>
<td><strong>Step 8:</strong> Presenting the canonical solution by contrasting it with the previous student solutions</td>
</tr>
<tr>
<td><strong>Canonical solution:</strong></td>
<td></td>
</tr>
</tbody>
</table>
For the third problem-solving step, together with the students the instructor identified how many pizza pieces a boy and a girl receives by drawing a circle representation of a pizza with $4/6$ (i.e., 4 pieces out of six pieces are marked) next to a circle representations with $2/3$ (i.e., 2 pieces out of three pieces are marked). While each boy receives more pieces, girls receive larger pieces. The instructor thus drew students’ attention to the relation-feature and the comparison-feature. In other words, as the fractions do not share the same denominator, students are not yet able to compare both amounts of pizza and to mathematically argue for the presence or absence of equivalence (cf. comparison-feature). In order to do so, the instructor highlighted the need to first find a common denominator (i.e., dividing both pizzas into the same amount of pieces) prior to actually comparing both amounts of pizzas with each other. As the common denominator of $2/3$ and $4/6$ is six, the instructor demonstrated how to expand the $2/3$ fractions. The instructor then counted the overall number of pieces (i.e., 6 pieces), marked four pieces out of the total of six pieces and asked students again whether boys or girls received a greater amount of pizza.

In order to enable students to practice the application of this deep feature (i.e., comparison-feature), they were asked to compare whether the birthday child’s aunts and uncles receive a smaller, greater or same amount of pizza (see Figures 3.4).

Note that in order to explore to which extent the developed instruction worked for the young target group (cf. first goal of study1) students’ answers to the practice activities concerning the division- and comparison-features were coded. In terms of both practice activities, students correct answers were rewarded with 1 point, whereas no or incorrect
answers were coded with 0 points. For the practice activity concerning the division-feature the correct answer was to equally and fairly divide the two pizzas amongst the three girls by making thirds, sixths and so on. For the practice activity concerning the comparison-feature the correct answer implied to demonstrate 1/4 equals 2/8 by for example expanding 1/4 or reducing 2/8.

<table>
<thead>
<tr>
<th>For three girls Steve ordered two “pizza Hawaii”. How can you equally and fairly divide the pizza amongst the girls? Please consider, you don’t need all given circles for solving the task</th>
<th>Steve also invited his aunts and uncles to his birthday party. He ordered pizza for them, too. They want to know whether a single aunt or uncle receives a greater amount of pizza when an aunt receives the blue marked piece of pizza and an uncle the two green marked pieces of pizza. Please compare and explain your answer.</th>
</tr>
</thead>
</table>

Figure 3.4. Activities to practice the division-feature and the comparison-feature.

**Group goal and role script**

In order to ensure fruitful collaborative processes during the problem-solving phase of PF and DI, the collaborating students were provided with a role script and a group goal. The role script consisted of two roles (i.e., role of thinker and questioner). While the student in the role of thinker was asked to explain his or her solution ideas to his or her learning partner, the student in the role of questioner was prompted to pose hint-questions and questions of clarification. During the collaborative problem-solving phase of the PF-Coll and DI-Coll conditions, students were provided with role cards displaying either the role of thinker or questioner and were asked to switch roles (at least once) (see appendix 7.1.2). Thus, students were prompted to collaborate following the role script, but were not forced to do so. This way, I aimed to prevent motivational loss due to over-scripting.

By additionally implementing a group goal, the aim was to give students a motivational incentive to indeed engage in team work and mutual support. The instructor announced there would be a prize (i.e., chocolate bars) for the dyad who best collaborated at the end of the study.
3.1.4 Instruments

Pretest

As it was not expected that the student sample would have any prior formal knowledge about the target domain, the pretest measured prerequisite mathematical knowledge rather than the target features underlying the equivalent fraction problem. Additionally, and in line with previous PF studies (e.g., Kapur, 2012; Loibl & Rummel, 2014a), by not measuring students’ knowledge about equivalent fractions it was also prevented to prompt students from all experimental conditions to generate solution ideas about the target features in the pretest. If students from all experimental conditions were prompted to generate solution ideas about the target features also the DI students were prompted to generate solution ideas prior to instruction which in turn would reduce differences between experimental conditions (e.g., Kapur, 2012; Loibl & Rummel, 2014a; Mazziotti, Loibl, & Rummel, 2015). Therefore, the pretest measured different mathematical knowledge prerequisites. For example, students were required to name fractions (first prerequisite) by showing students the fractional number 3/4 and asking them how they would label this number. A second mathematical prerequisite for being able to engage in problem solving was to divide natural numbers with and without remainder. For example, students were asked to divide 56 by 8 or 37 by 4. Finally, students’ initial understanding of fractions as part of a whole was tested by for example asking them to draw a picture representing 2/3. In total, students could reach nine points on the pretest.

Posttest

The conceptual knowledge posttest tested the aforementioned deep features of equivalent fractions. Overall, six items were administered that included 18 subtasks.

Story problem items were implemented that included graphical representations of fractions (mainly circles and rectangles) and a multiple choice item. Generally, students received one point for a correct answer and 0 points for no answer or an incorrect answer. The subtasks in which students were asked to compare fractions with each other were rewarded differently as these subtasks specifically required students to have understood the need to find a common denominator before actually comparing two or more fractions to each other (cf. comparison-feature). For these subtasks, students who were able to compare the given fractions with each other by referring to the canonical correct strategy of fraction expanding or reducing (in form of an explanation) received 2 points. Students
who applied a visual-based strategy by explaining to put pieces together received only 1 point. Students who did not compare the given fractions with each other received 0 points. The example item in Figure 3.5 illustrates these two different kinds of explanations and the way how points were awarded to these explanations. Overall, students could reach a maximum of 24 points on the posttest.

![Example item of the conceptual knowledge posttest in Study 1.](image)

### 3.1.5 Procedure

Inspired by the in vivo paradigm advocated by the Pittsburgh Science of Learning Center (Koedinger, Aleven, Roll, & Baker, 2009), Study 1 took place during students’ regular lessons. In the mathematics lesson preceding Study 1, all participants took the pretest (circa 10-15 minutes). At the beginning of the study, all students first received a 10-15 minutes introduction about the background and procedure of the study. Students in the PF-Coll and DI-Coll conditions further received explanations about the group goal and a brief role play illustrating productive use of the role script. To obtain insight into the collaborative problem-solving process, students’ collaboration in the PF-condition was videotaped. To do so, the problem-solving phase took place in a separate room next to the classroom. For the PF-Coll condition, pairs of students were taken out of the regular lessons one by one. After all pairs of the PF-Coll condition tried to solve the problem, they received instruction in their regular class room. In order to exclude influence from being videotaped on how students solved the problem, the students in the PF-Ind condition also tried to solve the PF problem in a separate room (in which the instructor and a video camera were present). Students were taken out of the classroom five at a time. It
was made sure students did not talk to each other but worked individually. Afterwards, all students within the PF-Ind condition received instruction in their classroom.

While students in the PF-Coll and PF-Ind conditions started with the problem-solving phase and then received instruction, students of the DI-Coll condition first received instruction and then solved an isomorphic problem (during which they were videotaped). The problem-solving phase took 30 minutes and the instruction phase took 45 minutes.

After both learning phases all students had 45 minutes to work on the conceptual knowledge posttest.

### 3.2 Results and discussion

This initial feasibility study had three goals: 1) to develop learning and test material, 2) to develop methods of analysis for detecting underlying learning mechanisms and 3) to preliminary test for the two main hypotheses (i.e., Hypotheses 1 and 2). This way, it was possible to derive implications for Study 2 and to develop methods which can be applied in Study 2. I will describe the findings and implications step-by-step.

#### 3.2.1 Findings and implications concerning study goal 1 and 2

Overall, the used learning material (i.e., the fraction problems and the instruction) and procedure (cf. study goal 1) worked well.

Figure 3.6 shows a typical student solution from a PF student. As one can see from the Figure, the students typically did not generate the canonical solution but solution ideas for each problem-solving step. Therefore, the problem at hand was complex for the young target group without being too challenging. For the students in the DI conditions the problem was also appropriate as students also generated solution ideas for each problem-solving step as documented on their solution sheets. In Study 2 students will thus again use the problems created in Study 1 (see Figures 3.2 and 3.3 in the method section of Study 1).

As the design of the instruction across all experimental conditions was identical and built upon the PF problem also the instruction ran rather smoothly and was understandable for the students. First indications for the effectiveness of the instruction can be derived from the extent to which the students correctly solved both practice activities that were handed out during instruction. As mentioned in chapter 3.1.3 the two practice activities asked
students for the two deep conceptual knowledge features, namely the division-feature and the comparison-feature. Across all experimental conditions 76% of students correctly solved the practice activity regarding the division-feature and 70% of students correctly solved the practice activity regarding the comparison-feature. Both relative numbers speak for a high understandability of the instruction. Therefore, this instruction will also be used in Study 2.

![Student solution example](image)

*Figure 3.6. Example for a typical student solution of the PF-problem (plus translation).*

The group goal and role script implemented to support students to engage in fruitful collaborative processes worked differently. Building upon the observational protocol taken during both learning phases and the test phase the group goal seemed to have had a relevance for students as they repeatedly asked at the beginning of the problem-solving phase and afterwards when they would get the group goal and what the group goal was about. Therefore, also in Study 2 this group goal will be implemented.
In contrast, by repeatedly watching the videos taken during the initial problem-solving phase of the PF-Coll condition it turned out some students at best adhered to their roles only during the initial “getting started phase” (ranging from minute 0 to approximately minute 5) by explicitly referring to the roles or by making use of the sentence opener illustrated on the role cards (see appendix 7.1.2). For example, in dyad A one student referred to his role card by explicitly saying “I am the thinker” in minute 4:31 whereas in dyad B the learning partners referred to their respective roles by explicitly saying “I am the thinker” and “I am the questioner” in minute 2:52 and minute 2:53 respectively. The students did not act as thinker or questioner during the remaining time of the problem-solving phase. For this, it seems as the role play only had a low effect on students’ collaboration. Due to the role play illustrating the use of the role script at the beginning of this study, the implemented role script probably rather had an indirect “sensitizing for fruitful collaboration” effect (if at all).

A possible explanation for the limited use of the role script could be that the young participants were not sufficiently familiar with applying a role script and with collaborative learning practices in general as such practices are often not well established in German classrooms and mathematics lessons at this young age. As the curricular standards for mathematics education (Ministerium für Schule und Weiterbildung des Landes Nordrhein-Westfalen, 2008) point out that even at this young age students should be able to collaborate and communicate their mathematical reasoning, this finding underlines the need to create more opportunities for young students to practice how to productively learn in a collaborative learning setting. Because practice to collaborate takes time (Westermann & Rummel, 2012) this is for sure no goal a researcher can reach during the limited amount of study time scheduled for in vivo school experimentations by for example establishing a week-long collaborative training. However, as students’ fruitful collaboration during problem solving is such an important component of the thesis at hand (cf. Hypothesis 2, 3, 4), the question is which design implication can be derived for Study 2. Given the time constraints the main implication for the design of Study 2 is to recruit students who ought to be more capable to collaborate as they should already have had more opportunities to practice to collaborate with a peer. Simultaneously, these students should still meet the PF requirement of not having any prior formal knowledge about the target domain (here about equivalent fractions). It is thus planned to recruit German fifth graders rather than German fourth graders for Study 2 who ought to have had more opportunities to practice to collaborate. According to the German standards of the core curriculum
(Ministerium für Schule und Weiterbildung des Landes Nordrhein-Westfalen, 2007; Ministerium für Schule, Jugend und Kinder des Landes Nordrhein-Westfalen, 2004) and in consultation with potential schools the fifth grader share with the fourth graders to not have any formal prior knowledge about the target domain but similar to the fourth graders these students already have developed pre-concepts about equivalent fractions. In addition, as the implemented role script due to the role play probably had a “sensitizing for fruitful collaboration” effect the implication for Study 2 was to continue to use the role script while lowering the expectation about its effectiveness.

With view on the procedure and scheduled time the observational protocols further elicited the scheduled time for both learning phases was appropriate. More precisely, by repeatedly watching the videos and annotating the time (of solution idea generation) it turned out PF students had sufficient time to generate more than three single solution ideas for the three problem-solving steps. Simultaneously, at the end of this learning phase the young students did not continue to generate further solution ideas after repeated motivational prompting as the observation protocol and the videos indicated. Nevertheless, based on informal discussions with participating teachers and on students’ feedback concerning the overall study experience at the end of Study 1 the fourth graders were quite exhausted after both learning phases and the conceptual knowledge post test phase. This finding further underlined the need to recruit fifths graders rather than fourth graders who have slightly higher cognitive capacities (Gathercole et al., 2004) and should thus be more capable to withstand the study demands.

A further design implication for Study 2 concerning the procedure was related to the following: To be able to take videos of single dyads (and individually learning students) in PF the problem-solving phase of Study 1 took place in a separate room and students were taken out one by one prior to receive the instruction with all their class mates in the class room. This way, it was possible to explore how students worked on the equivalent fraction problem and to ensure to make videos in appropriate quality which was required to reach the second goal of Study 1, namely to develop a coding scheme for capturing different types of collaborative processes (i.e., verbal utterances). However, as this goal is reached by the end of Study 1 there is no longer a need for taking single dyads or individually learning groups of students out during problem-solving in Study 2. With view on Study 2 I thus derive the implication to let all learning phases take place in the class room (leading to slight changes in the procedure). A further advantage of conducting all learning phases in a (single) class room was to increase the external validity of Study 2.
While the learning materials and collaborative support structure worked quite well, the conceptual knowledge posttest worked only to a much lower extent. As shown in Table 3.4 illustrated in chapter 3.2.3, students acquired a rather low number of points ranging from 5 points to 8 points whereas the maximum number of possible points lies at 24 points. This speaks for floor effects which points at the need to revise the posttest for Study 2. In particular, I aim to lower the level of difficulty for the conceptual knowledge posttest in Study 2.

Summing up, I derive the conclusion regarding study goal 1 to basically keep the learning material, the procedure and the implemented collaborative support structure while I plan to revise the conceptual knowledge posttest and to change the recruiting strategy.

In order to be able to detect the three aforementioned learning mechanisms underlying PF and to test for Hypotheses 3 and 4 it is required to develop methods of analysis that asses process-related measures such as the number of student-generated solution ideas (cf. learning mechanisms 1), the solution quality (cf. learning mechanism 2) and students’ collaborative processes (cf. learning mechanism 1 and 2). For this, the number of student-generated solution ideas, the solution quality and students’ collaborative processes of this study were used as resources for developing methods of analysis. These methods in turn are applied to the respective data collected in Study 2.

To develop a coding method for the number of student-generated solution ideas detecting learning mechanism 1 (i.e., students’ activation and differentiation of prior knowledge) and to develop a coding method for the solution quality detecting the existence or absence of students’ knowledge gaps paving the way for learning mechanism 2, I built upon the solution sheets students used during the problem-solving phase. As shown in Figure 3.6, students documented their solution ideas on this sheet. As the coding of the number of student-generated solutions and solution quality will be applied to the data collected in Study 2, I will report in more detail about the way students’ solutions are coded in chapter 4.1.4.

In order to be able to assess different types of collaborative processes paving the way to investigate how these different collaborative processes affect the evolvement of the aforementioned learning mechanisms and students’ conceptual knowledge in turn an additional coding method (i.e., coding scheme) needed to be developed. For this, I followed in line with Meier and colleagues (Meier, Spada, & Rummel, 2007) a top-down and bottom-up
approach. This means, the coding scheme needed to be rooted in theory about collaborative learning and empirical data reflecting the particularities of the target group (i.e., young students’ verbal utterances). A first step towards the development of this kind of coding scheme was thus to transcribe the videotaped dialogues of the eight dyads trying to solve the equivalent fraction problem in the PF-Coll condition of this study (bottom-up perspective). The final coding scheme and the procedure of its application are described in chapter 4.1.4.

Summing up, this initial feasibility study also reached its second goal to serve as a resource for developing methods of analysis for detecting the three aforementioned learning mechanisms through analyzing process-related measures. This way, it will be possible to examine the extent to which students activate and differentiate their prior knowledge (cf. learning mechanisms 1), become aware of knowledge gaps (cf. learning mechanism 2) and recognize deep target features (cf. learning mechanisms 3) in Study 2.

3.2.2 Preliminary results and discussion concerning the main hypotheses (study goal 3)

The third goal of Study 1 was to serve as first test of the two main hypotheses (i.e., Hypothesis 1 and 2). Due to the small sample size of 52 students resulting in low statistical power and the alignment of a single class to a single conditions the following results must be treated with care.

Prior analyses

Means and standard deviation of students’ pretest scores are displayed in Table 3.3. Students from all three experimental conditions seem to have started from an approximately equal prior knowledge starting point as the one-factorial ANOVA with pretest as dependent variable and condition as factor revealed ($F[2,49] = 1.56, p=.22$). Students’ prior knowledge as explained above did not include the target concepts and thus does not correlate with students’ conceptual knowledge at posttest ($r(50) = .03 p=.83$). For this, it was not included as covariate for further analyses.

In order to address the issue of different sample sizes between the experimental conditions, I calculated a Levene test for homogeneity of variances and found the differences between variances not to be statistically significant, $F[2, 49]=1.44, p=.25$. 

Table 3.3. Mean and standard deviations of pretest scores of Study 1.

<table>
<thead>
<tr>
<th></th>
<th>$M$</th>
<th>$SD$</th>
<th>$N$</th>
</tr>
</thead>
<tbody>
<tr>
<td>PF-Coll</td>
<td>4.50</td>
<td>2.63</td>
<td>16</td>
</tr>
<tr>
<td>PF-Ind</td>
<td>5.12</td>
<td>1.90</td>
<td>17</td>
</tr>
<tr>
<td>DI-Coll</td>
<td>5.68</td>
<td>1.29</td>
<td>19</td>
</tr>
</tbody>
</table>

Students’ conceptual posttest

In order to preliminary test for Hypothesis 1 and 2 a one-factorial ANOVA with students’ conceptual knowledge as dependent variable and experimental condition as independent variable was calculated (see also Mazziotti et al., 2015; Mazziotti, Rummel, & Loibl, 2015). Means and standard deviations of students’ conceptual knowledge posttest scores are displayed in Table 3.4. There were no significant differences between experimental conditions ($F[2, 49]=2.12, p=.13$).

Table 3.4. Means and standard deviations of students' conceptual knowledge at posttest of Study 1.

<table>
<thead>
<tr>
<th></th>
<th>$M$</th>
<th>$SD$</th>
<th>$N$</th>
</tr>
</thead>
<tbody>
<tr>
<td>PF-Coll</td>
<td>7.88</td>
<td>5.43</td>
<td>16</td>
</tr>
<tr>
<td>PF-Ind</td>
<td>7.71</td>
<td>3.29</td>
<td>17</td>
</tr>
<tr>
<td>DI-Coll</td>
<td>5.26</td>
<td>3.91</td>
<td>19</td>
</tr>
</tbody>
</table>

In order to separately analyze the effect of timing of instruction with instructor-led comparing and contrasting activity and students’ collaboration during the initial problem-solving phase of PF two a priori contrasts were defined. In order to further explore the effect of the timing of instruction (with comparing and contrasting activity) on students’ conceptual knowledge acquisition I compared both PF conditions (PF-Coll and PF-Ind) with the DI-Coll condition. In line with Hypothesis 1, I found a moderate and significant effect of timing of instruction ($F[1, 49]=4.23, p=.05, \eta^2 = .08$). Students in both PF-conditions acquired around two points more on the conceptual knowledge posttest as compared to students within the DI-Coll condition. Contrary to Hypothesis 2, the second a priori contrast comparing the PF-Coll condition with the PF-Ind condition did not reveal
a significant difference despite the descriptive superiority of the PF-Coll condition over PF-Ind condition, \(F[1, 49]=.01, p=.91\).

Against this background I would like to discuss these preliminary results and derive further implications for Study 2.

As Hypothesis 1 could be confirmed it seems as delaying instruction after problem-solving worked for young students. Because in this study the instruction includes the typical instructor-led comparing and contrasting activity it seems as learning mechanisms 2 (i.e., students’ awareness of failure) and learning mechanism 3 (i.e., students’ recognition of deep target features) came to play (cf. Loibl et al., 2016). Due to the cascade effect (Loibl et al., 2016) I might derive the cautious conclusion that also learning mechanism 1, namely students prior knowledge activation and differentiation during problem solving, preceding the aforementioned learning mechanisms 2 and 3 evolved. However, due to the low sample size and the low number of points students reached in the conceptual knowledge posttest (24 points maximum; maybe pointing to floor effects) I would rather treat the results of Study 1 with caution and as preliminary. For Study 2, I thus derive the implication to increase the sample size.

With regard to Hypothesis 2, it seems as students did benefit only to a very low extent from collaborating with a peer. Interestingly, students from the PF-Coll condition reached very high standard deviations at their conceptual knowledge posttest which in turn might point towards great differences between the ways how students actually collaborated during problem solving. As in light of the interaction paradigm (Dillenbourg et al., 1996) different kinds of collaborative processes have differential effects on students’ learning outcome (here: conceptual knowledge) and probably on the evolvement of three core learning mechanisms underlying PF the need to conduct intensive collaborative process analyses (in the next study) is once again emphasized.

Summing up, by developing learning and test material for a young target group, by developing methods of analysis for detecting underlying learning mechanisms and by finding first indications for the two main hypotheses, Study 1 reached all its goals.
4 Study 2

Building upon the results of Study 1 the goals of Study 2 were to shed more light on the effectiveness of delaying instruction after problem solving with improved study design and to systematically test for all hypotheses.

To reach these goals a much larger sample size (N= 228) was recruited and a complete 2x2 experimental set up that systematically varied the factors timing of instruction with instructor-led comparing and contrasting activity (problem solving prior to instruction, PF vs. problem solving after instruction, DI) and social form of learning (collaborative learning during problem solving vs. individual learning during problem solving) was installed.

In a first step Study 2 tested for the two main hypotheses: Students learning with PF acquire more conceptual knowledge as compared to students learning with DI (including an instructor-led comparing and contrasting activities) (cf. Hypothesis 1). Students trying to collaboratively (i.e., in a dyad) solve an equivalent fraction problem acquire more conceptual knowledge as students trying to individually solve an equivalent fraction problem (cf. Hypothesis 2). In order to test for Hypothesis 1 both PF conditions were compared with both DI conditions. In order to test for Hypothesis 2 I compared collaboratively learning students with individually learning students trying to solve the equivalent fraction problem prior to receive instruction (i.e., PF-Coll and PF-Ind conditions).

To additionally test for Hypotheses 3 and 4 and to investigate the extent to which the beneficial learning mechanisms underlying PF evolved a second strong focus of Study 2 was to conduct intensive process analyses by relying on multiple sources of data.

In order to investigate learning mechanism 1, namely students’ prior knowledge activation and differentiation, the number of student-generated solution ideas during the initial problem-solving phase of PF was analyzed. As in some previous PF studies with older samples the number of student-generated solution ideas has been connected to students’ learning outcome, the link between students’ solution ideas and the students’ conceptual knowledge was additionally examined. As students’ collaboration is a central PF design component that specifically addresses students’ prior knowledge activation and differentiation, I further tested whether the number of student-generated solutions differs across both PF-conditions. In line with Hypothesis 3, I expect collaboratively learning students to generate a higher number of solution ideas than individually learning students. In light
of the interaction paradigm it was examined how different kinds of collaborative processes are linked to the number of solution ideas and to students’ conceptual knowledge acquisition. As the analyses of different collaborative processes builds upon students’ verbal utterances, in the following I will refer to collaborative utterances to label this kind of processes.

In order to evaluate to which extent learning mechanism 2, namely students’ awareness of knowledge gaps, came to play I relied on the quality of students’ solutions and how the quality is linked to students’ conceptual knowledge. The solution quality to some extent reflects the existence of gaps in students’ knowledge and thus served as a proxy for students becoming aware of these gaps. In line with Hypothesis 4, it was examined whether students in the PF-Coll condition generate qualitatively better solution as compared to students in the PF-Ind condition. Also for solution quality it was investigated how different kinds of collaborative utterances are linked to the solution quality.

In order to investigate the extent to which learning mechanism 3, namely students’ recognition of deep features of the target domain, evolved I relied on the conceptual knowledge posttest as it specifically requires students to recognize and understand deep features of the target domain. No additional tests asking students for deep features of the target domain during the learning phases were implemented. This way, (especially) the instruction phase was kept as similar to typical PF implementations as possible and it was ensured to keep the overall learning experience of Study 2 on an appropriate and thus not overwhelming level for the young target group.

### 4.1 Method

In this study the same learning material, collaborative support structure (i.e., group goal and role script) and pretest were used as in Study 1 (see method chapters 3.1.3 and 3.1.4). As already mentioned in the discussion of Study 1, for Study 2 the following aspects were changed:

In order to increase power, a much larger sample was recruited and the two factors, namely the timing of instruction with instructor comparing and contrasting activity and the social form of learning, were varied in a complete 2x2 study design. To ensure students to be a bit more experienced with learning in a collaborative setting and to prevent students from being overly exhausted by the end of Study 2 or already at the conceptual knowledge posttest fifth graders rather than fourth graders were recruited. To lower the
difficulty level of the conceptual knowledge posttest I again collaborated with mathematics education teachers and designed a conceptual knowledge posttest asking for the three equivalent fraction features (i.e., relation-, division- and comparison-features). In order to increase external validity this time all learning (and test) phases took place in the classroom. This means, in contrast to Study 1 no single dyads or groups of individually learning students were taken out in a separate classroom to engage in problem-solving one after the other. They all learned simultaneously in the same classroom and were audiotaped by a dictation machine. As this was the only procedural change of Study 2 as compared to Study 1 the procedure of Study 2 is detailed in chapter 3.1.3 (under procedure).

For the purpose of clarity, I will first report in more detail about the methods being adapted from Study 1. In addition, as Study 2 aims to investigate how all three learning mechanisms underlying PF evolved I will further describe the methods of analysis for these analyses in chapter 4.1.4.

4.1.1 Design

In order to test the aforementioned hypotheses and to analyze the extent to which the aforementioned learning mechanisms come to play (i.e., processes analyses) a quasi-experimental study with a complete 2x2 design including the factors timing of instruction with instructor-led comparing and contrasting activity (i.e., problem solving prior to instruction, PF vs. problem solving after instruction, DI) and social form of learning (i.e., collaborative learning during the initial problem-solving phase vs. individual learning during the initial problem-solving phase) was conducted. In both PF-conditions students either tried to collaboratively (i.e., in a dyad in the PF-Coll condition) or to individually (i.e., PF-Ind condition) solve a yet unknown equivalent fraction problem illustrated in Figure 3.2. In both DI-conditions students first received instruction and then either worked collaboratively (i.e., DI-Coll condition) or individually (i.e., DI-Ind condition) on an equivalent fraction problem (see Figure 3.3). Table 4.1 provides an overview about the four experimental conditions and the sequence of the respective learning phases. Nine classes were recruited from three different schools and were randomly assigned to experimental conditions. In doing so, it was ensured to not have more than a single class from a school in an experimental condition.
Table 4.1. Overview about experimental conditions and the learning phases in Study 2

<table>
<thead>
<tr>
<th>Timing of instruction</th>
<th>Social form of learning</th>
<th>Name of condition</th>
<th>Learning phase 1</th>
<th>Learning phase 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>collaborative</td>
<td>PF-Coll</td>
<td>Collaborative problem solving</td>
<td>Instruction</td>
</tr>
<tr>
<td></td>
<td>individual</td>
<td>PF-Ind</td>
<td>Individual problem solving</td>
<td>Instruction</td>
</tr>
<tr>
<td>Direct Instruction</td>
<td>collaborative</td>
<td>DI-Coll</td>
<td>Instruction</td>
<td>Collaborative problem solving</td>
</tr>
<tr>
<td>(DI)</td>
<td>individual</td>
<td>DI-Ind</td>
<td>Instruction</td>
<td>Individual problem solving</td>
</tr>
</tbody>
</table>

4.1.2 Sample

In line with the discussion of Study 1 (see chapter 3.2.1), for Study 2 228 German fifth graders, who are usually between 10 and 12 years old, were recruited. The fifth graders (should) have slightly higher cognitive capacities as compared to the fourth grade students (Gathercole et al., 2004). Note that in Germany the fifths graders have just finished elementary school which lasts from grade 1 to grade 4. Across all experimental conditions parental consent for students’ involvement was obtained for all participating students.

4.1.3 Posttest

As in Study 1 and other PF studies the posttest used for Study 2 asked for students’ conceptual knowledge. As in Study 1 also this conceptual knowledge posttest asks students for the aforementioned deep features of equivalent fractions, namely the relation-feature, the division-feature and the comparison-feature (for more details see chapter 3.1.3). Because students of Study 1 achieved only a very low number of points at the conceptual knowledge I adapted the level of difficulty for the conceptual knowledge posttest of this study.

Overall, I administered six items including several subtasks. I again implemented story problem items that included graphical representations of fractions (i.e., mainly circles and number lines) and asked students for an explanation-based answer rather than “just” a numerical answer. This way, it was possible to identify whether students actually understood the aforementioned deep features. In addition to the story problem items I included an item in a multiple choice format.
In order to code students’ responses of the story problem items two categories of explanations were distinguished (similar as in Study 1): The canonically correct explanation (e.g., comparing two fractions by expanding or reducing one the fractions for reaching the same denominator and numerator for both fractions) and an explanation based on a visual strategy (e.g., comparing two fractions by visually putting smaller pieces of fraction A together to get the same size of one piece of fraction B). I rewarded the canonically correct explanation with 2 points, the visual based explanation with 1 point and any other (or no) explanation with 0 points. Figure 4.1 shows an example for both categories of explanations. For the multiple-choice item students were rewarded with 1 point for the correct answer and an additional point when they were able to correctly visualize their answer. This way, I ensured also with this format that students understood the answer they had chosen and thus reduced the probability of guessing.

Figure 4.1. Example item of the conceptual knowledge posttest in Study 2.

In total students could reach 18 points. Given the low number of posttest items and the measurement of three different deep conceptual features the internal consistency of the posttest was satisfactory ($\alpha = .54, n= 228$) as alpha assumes to measure one global one-dimensional construct rather than three different features (Bortz & Döring, 2006, p. 198).

### 4.1.4 Process analyses to detect underlying learning mechanisms

The process analyses focused on the extent to which the three core learning mechanisms underlying PF came to play by referring to the number of student-generated solution
ideas, to the solution quality and to different kinds of collaborative utterances. In this context, also Hypothesis 3 and 4 will be tested. The coding methods for all process-related measures are an outcome of Study 1 as they were developed by making use of the respective data collected in Study 1 (cf. second goal of Study 1).

**Number of solution ideas**

In order to investigate whether *learning mechanism 1*, namely students’ prior knowledge activation and differentiation worked, I coded and analyzed students’ solutions with regard to the number of solution ideas. Naturally, the number of solution ideas is closely linked to the equivalent fraction problem and its three problem-solving steps (for more details see Figures 3.2). To determine the number of students’ solution ideas I used a fine-grained approach as I coded the number of solution ideas per problem-solving step rather than the number of complete solutions (i.e., students attempt to engage in *all three* problem-solving steps) as unit of analysis. As a coding example, drawing the four pizza Salami and dividing them for the six boys into halves was coded as one solution idea for problem-solving step 1, and drawing and dividing the four pizza Salami for the six boys into quarters was coded as another solution idea for problem-solving step 1. Because coming up with a new solution idea for a single problem-solving step is less challenging than coming up with three new solution ideas for all three problem-solving steps (i.e., complete solution) coding on the level of problem-solving ideas was more appropriate for the young target group than coding on the level of a complete solution. I summed up the number of solution ideas across all problem-solving steps into one sum score.

Note a preparatory step for developing a coding method assessing the number of solution ideas was to align boxes around all solution ideas and to label them as for example solution idea 1 for problem-solving step 1. This way, the nature of the data was addressed and the likelihood of ambiguity decreased.

Two coders were intensively trained to assess the number of solution ideas (and the solution quality). To determine interrater reliability for the number of solution ideas, the raters both coded 43% of the solution sheets with satisfactory interrater reliability (number of solution ideas $\kappa = .65$). Disagreements were resolved by discussion.

**Quality of students’ solutions**

For exploring to which extent students had gaps in their knowledge and could thus become aware of these gaps (cf. *learning mechanisms 2*) the solution quality was analyzed.
For the solution quality, it was rated how closely the student-generated solutions resembled the canonical solution. Indeed, it was rated to which extent students were able to equally and fairly divide the two amounts of pizzas for boys and girls respectively (cf. problem-solving step 1 and 2; division-feature) and to which extent they were able to compare both amounts of pizzas with each other (cf. problem-solving step 3; comparison-feature). Table 4.2 illustrates how many points students could receive for each problem-solving step. As unit of analysis a single solution was used, that is, students’ attempt to engage in all three problem-solving steps before restarting a new solution by engaging again in problem-solving step 1. Because 76% of the students across both PF-conditions generated only one complete solution (i.e. one attempt at solving all three problem-solving steps), I only rated the quality of this first solution. Only 24% of students generated a second solution. And further: Only a very small amount of students (i.e. 11% = 13 students out of a total of 117 students) engaged in a second complete solution (i.e., all three problem-solving steps). In sum, assessing the quality of all problem-solving steps leads to a maximum possible score of 9.

To determine interrater agreement for the solution quality, the two trained raters both rated a subset of 20% of students’ first solution. Agreement was high (ICC_{absolute} = .79; 95%-CI [.53; .92]). Again disagreements were resolved by discussion.
Table 4.2. Rating of the solution quality with examples

<table>
<thead>
<tr>
<th>Problem-solving step</th>
<th>Basic rationale of the rating</th>
<th>Examples for student solutions</th>
<th>Explanation of the rating for the given example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dividing the pizza for boys and girls</td>
<td>Dividing equally, fairly and without a remainder (e.g., into sixths, thirds, twelfths). (1 point for trying to divide + 1 point for dividing equally + 1 point for dividing without leaving a remainder = 3 points)</td>
<td><img src="image1.png" alt="Image" /></td>
<td>1 point for trying to divide + 1 point for dividing equally (i.e., through the central points of the circles) + 1 point for dividing without leaving a remainder (i.e., into sixth leading two 24 divided by 6 equals 4)</td>
</tr>
<tr>
<td></td>
<td>Diving equally but unfairly (e.g., halves, quarters, fifths). (1 point for trying to divide + 1 point for dividing equally = 2 points)</td>
<td><img src="image2.png" alt="Image" /></td>
<td>1 point for trying to divide + 1 point for dividing equally (i.e., through the central point) 0 points for leaving a remainder as 32 divided by 6 equals 5 with remainder 2</td>
</tr>
<tr>
<td></td>
<td>Avoidance strategies: Stating that some children are not hungry, or ordering more pizza so that children get whole pizzas. (1 point for trying to divide whole pizzas amongst the children = 1 point)</td>
<td><img src="image3.png" alt="Image" /></td>
<td>1 point for trying to divide whole pizzas amongst the children Because the example solution considers an additional person (i.e., Steve, the birthday child) this solution reflects an avoidance strategy</td>
</tr>
<tr>
<td></td>
<td>Not trying to divide (0 points)</td>
<td><img src="image4.png" alt="Image" /></td>
<td></td>
</tr>
<tr>
<td>Problem-solving step 3: Comparing two amounts of pizza</td>
<td>Both pizzas are divided into the same amount of pieces (e.g., both pizza Salami and pizza Hawaii are divided into twelfths). Students can thus “prove” that each boy and girl receives the same amount of pizza. (1 point for trying to compare + 1 point for comparing the equal amounts of pieces + 1 point for reaching the correct conclusion that each child receives the same number of pieces = 3 points)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>The two pizzas are divided into different numbers of pieces (e.g., pizza Salami into thirds and pizza Hawaii into twelfths). Due to a superficial comparison, students conclude correctly that each boy and girl receives the same amount of pizza. (1 point for trying to compare + 1 point for reaching the correct conclusion that each child receives the same number of pieces = 2 points)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Translation: Both receive the same as we cut all pizza into quarters except for the last, which we have cut into fifths.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Translation: 6 boys: 4 pizza each 6 pieces 3 girls: 2 pizza each 6 pieces A.: Everybody gets the same amount</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1 point for trying to compare + 1 point for comparing the equal amounts of pieces (i.e., sixth for boys and girls) + 1 point for reaching the correct conclusion that each child receives the same number of pieces</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1 point for trying to compare + 0 point for comparing the unequal amounts of pieces (i.e., quarters and fifths) + 1 point for reaching the correct conclusion that each child receives the same number of pieces</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The two pizzas are divided into different numbers of pieces (e.g., pizza Salami into thirds and pizza Hawaii into eights). Due to a superficial comparison, students conclude incorrectly that either a boy or a girl receives a smaller or greater amount. (1 point for trying to compare = 1 point)

Translation: The boys get more pieces of pizza!!! The boys get eight pieces of pizza!!!

| Not trying to compare (0 points) | 1 point for trying to compare + 0 point for comparing the unequal amounts of pieces (i.e., eights for boys and quarters for girls) + 0 point for reaching the incorrect conclusion that boys receive more pizza |
Coding and analyzing different types of collaborative utterances

As explained above, in order to be able to examine how different types of collaborative utterances affect the evolvement of the aforementioned learning mechanisms underlying PF and students’ conceptual knowledge in turn a coding scheme\(^1\) was in line with a top-down and bottom-up approach (Meier et al., 2007) developed. The coding scheme is on the one hand rooted in the ICAP-framework (Chi, 2009; Chi & Wylie, 2014) and broader literature about collaborative communication processes (e.g., Clark, Herbert, H. & Brennan, 1991) and literature about students’ off-task behavior (e.g., Godwin et al., 2016). On the other hand, the coding scheme builds upon verbal process data from Study 1 in which dyads tried to solve the very same equivalent fraction problem prior to receiving instruction (cf. study goal 2 of Study 1). By following a top-down and bottom-up approach a mapping between the particularities of the actual collaborative interaction data (i.e., young students’ verbal utterances) and different theoretical dimensions about students’ collaborative learning was ensured. The coding scheme was developed over several design loops and iteratively improved.

As a preparatory step for coding and analyzing the verbal interaction data (i.e., audiotaped utterances) collected during the problem-solving phase of the PF-Coll condition (of Study 2) audiotapes of 34 dyads were transcribed. The audio data of three dyads were lost due to technical issues. In a first step, the transcripts were segmented into single turns of a student within a dyad. This way, it was possible to ensure a high degree of objectivity as this unit of analysis leaves no space for interpretation. In order to examine to which extent different collaborative utterances occurred during the initial problem-solving phase of PF and how they are linked to the number of solution ideas (cf. learning mechanism 1), the solution quality (cf. learning mechanism 2) and to students’ conceptual acquisition the coding scheme covers and differentiates between six kinds of collaborative processes in total: interactive, constructive, active, coordinative, back channeling and off-task collaborative utterances. While the labels interactive, constructive and active utterances are derived from the ICAP-framework, the labels coordinative, back channeling and off-task utterances are taken from the broader literature about collaborative learning and students’ off-task processes. Note, from the ICAP-framework passive processes were left out as this type of learning processes does not imply any cognitive engagement of the learner.

\(^1\) I collaborated with Dr. Anne Deiglmayr from the ETH Zurich and with my supervisor Prof. Nikol Rummel to develop this coding scheme.
and was thus out of the research interests. Table 4.3 provides an overview about all six codes. Codes were aligned to single utterances. An utterance was identified to be *interactive* if the utterance of one learning partner builds upon his learning partner’s prior utterance (e.g., solution idea). For example, learning partner A explains, discusses, rephrases, and describes a solution idea of his learning partner B. In addition, learning partner A’s utterance has to reflect a mathematical content and this mathematical content has to go beyond the information included in the equivalent fraction problem (see learning material section) for labelling his utterance as interactive. In order to assign the code *constructive* to a students’ utterance it can reflect a self-explanation, a description or a paraphrase of his or her solution idea. In parallel to interactive utterances also constructive utterances need to reflect a mathematical content which goes beyond the information presented in the equivalent fraction problem. There is only one decisive difference between constructive and interactive utterances: While interactive utterances take up the learning partner’s solution idea by building upon his or her prior utterance, constructive utterances do not (or at least not explicitly) take into account what the learning partner has previously said. An utterance was labeled as *active* when students read aloud the problem at hand or parts of it and thus paid attention to the relevant mathematical content. If a student’s utterance included suggestions for next problem-solving steps (e.g., let’s start with problem-solving step 2), a review of problem-solving steps students have already undertaken (e.g., we have already done problem-solving step 1) or a management of turn takings (e.g., “you go first, it is your turn”) and did not reflect a mathematical content going beyond the presented learning material then I coded this utterance as *coordinative*. An utterance was labeled as back channeling when students signaled they agreed or disagreed with the previous utterances (by saying something like yes, ok, I understand), signaled to continue to listen to the learning partner and reinsured the learning partner understood what they were saying (e.g., did you understand?). Therefore, it does not cover any mathematical content. Finally, if students talked about anything else but the problem at hand by referring to the learning partner’s new haircut, the weather or anything else this utterance was labelled as *off-task* collaborative utterance. Overall, I used a very conservative coding approach as I decided for the lower code in case of ambiguity.

In order to be able to apply the coding scheme two coders were intensively trained by making use of a series of anchor examples and borderline cases for each utterance label (i.e., interactive, constructive, active, coordinative, back channeling and off-task). Inter-
rater agreement was tested with a subset of 15% of randomly selected transcripts. Interrater agreement across all codes was high (ICC$_{\text{absolute}} = 0.92$; 95%-CI [.49; .96]). Inconsistencies were resolved by discussion.

In order to analyze how different kinds students’ collaborative utterances are associated to the number of solution ideas (cf. learning mechanism 1), the solution quality (cf. learning mechanism 2) and students’ conceptual knowledge the sum (i.e., absolute frequencies) of individual students’ utterances labelled as interactive, constructive, active, coordinative, back channeling or off-task respectively was calculated.
Table 4.3. Overview of coding scheme with coding examples

<table>
<thead>
<tr>
<th>Codes</th>
<th>Conditions for labeling an utterance as</th>
<th>Examples (translated by the author)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interactive</td>
<td>• Builds upon a prior contribution of the learning partner</td>
<td>Student 1: Well, when we divide the pizza for boys into halves, then the first boy gets a half, the second gets a half…</td>
</tr>
<tr>
<td></td>
<td>• Reflects a mathematical content</td>
<td>Student 2: …the fourth gets a half, the fifth gets an entire and the sixths gets an entire piece.</td>
</tr>
<tr>
<td></td>
<td>• Mathematical content reflects an idea that goes beyond the information included in the equivalent fraction problem*</td>
<td></td>
</tr>
<tr>
<td>Constructive</td>
<td>• Does not build upon a prior contribution of the learning partner</td>
<td>Student 3: Well, this is six times four, four pizzas and then we can divide six by four and four…hmh, four times six are 24.</td>
</tr>
<tr>
<td></td>
<td>• Reflects a mathematical content</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Mathematical content reflects an idea that goes beyond the information included in the equivalent fraction problem*</td>
<td></td>
</tr>
<tr>
<td>Active</td>
<td>• Students read aloud the equivalent fraction problem or parts of the problem</td>
<td>Student 2: For the six boys he orders four pizza Salami and for the three girls he orders two pizza Hawaii, because the boys share the pizza amongst each other. Also the girls share the pizza amongst each other.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Student 6: Yes, look. Who receives more? A boy or a girl receives a greater proportion of pizza.</td>
</tr>
<tr>
<td>Coordinating</td>
<td>• Does not reflect a mathematical content that goes beyond the information included in the equivalent fraction problem</td>
<td>Student 4: I have to, hmm, well we have already painted but we did not calculate yet.</td>
</tr>
<tr>
<td>Back channeling</td>
<td>Student 7: Hmhm, yes.</td>
<td></td>
</tr>
<tr>
<td>-----------------</td>
<td>-----------------------</td>
<td></td>
</tr>
<tr>
<td><strong>•</strong> Does not reflect any mathematical content</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>•</strong> Reflects agreement or disagreement</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>•</strong> Indicates the learning partner to continue to listen</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>•</strong> Reinsures the learning partner’s understanding</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Off-Task</th>
<th>Student 8: Did you understand?</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>•</strong> Any other utterance that does not refer to the equivalent fraction problem</td>
<td></td>
</tr>
<tr>
<td><strong>•</strong> E.g., talking about the weather or a new haircut</td>
<td></td>
</tr>
</tbody>
</table>

*Examples for mathematical content that goes beyond the presented material are dividing, adding, multiplying fractions or using graphical representations showing fractions*
4.2 Results

4.2.1 Prior analyses

Means and standard deviations of the pretest scores are shown in Table 4.4. When calculating an one-way ANOVA with the factor condition and pretest as dependent variable it turned out students across all experimental conditions begun from an approximately equal prior knowledge starting point ($F[3,224]=1.01, p=.39$). As the pretest score was not correlated with students’ conceptual knowledge in the posttest ($r(226)=.01$, $p=.15$), it was not included as a covariate in further analyses (Loibl & Rummel, 2014a).

Table 4.4. Mean and standard deviations of pretest scores of Study 2

<table>
<thead>
<tr>
<th>Condition</th>
<th>$M$</th>
<th>$SD$</th>
<th>$N$</th>
</tr>
</thead>
<tbody>
<tr>
<td>PF-Coll</td>
<td>4.19</td>
<td>2.48</td>
<td>74</td>
</tr>
<tr>
<td>PF-Ind</td>
<td>4.00</td>
<td>2.04</td>
<td>49</td>
</tr>
<tr>
<td>DI-Coll</td>
<td>4.74</td>
<td>2.25</td>
<td>57</td>
</tr>
<tr>
<td>DI-Ind</td>
<td>4.40</td>
<td>2.46</td>
<td>48</td>
</tr>
</tbody>
</table>

Due to the different sample sizes in each condition a Levene test of homogeneity of variance was calculated and showed no statistically significant differences between variances with view on students’ conceptual knowledge, ($F[3,224]=.73, p=.53$).

4.2.2 Conceptual knowledge posttest score

Table 4.5 shows the means and standard deviations of students’ conceptual knowledge posttest score. To test Hypotheses 1 and 2, I calculated a two-factorial ANOVA with the timing of instruction (PF vs. DI) and the social form of learning (collaborative vs. individual) as factors, and conceptual knowledge as dependent variable (see also Mazziotti, Loibl et al., 2015). The ANOVA revealed neither a significant main effect of timing of instruction ($F[1,224]= 1.01, p=.32$) nor of the social form of learning ($F[1,224]= 3.36, p=.06$) nor a significant interaction effect, ($F[1,224]=.18, p=.67$). As the missing main effect of timing shows, there was no advantage of problem-solving prior to instruction (i.e., PF). Hypothesis 1 thus receives no support.
Table 4.5 Means and standard deviations of students’ conceptual posttest scores at post-test of Study 2

<table>
<thead>
<tr>
<th>Timing of instruction</th>
<th>Social form of learning</th>
<th>Posttest scores</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>M</td>
</tr>
<tr>
<td>PF</td>
<td>collaborative</td>
<td>10.31</td>
</tr>
<tr>
<td></td>
<td>individual</td>
<td>10.80</td>
</tr>
<tr>
<td></td>
<td>total</td>
<td>10.51</td>
</tr>
<tr>
<td>DI</td>
<td>collaborative</td>
<td>10.50</td>
</tr>
<tr>
<td></td>
<td>individual</td>
<td>11.28</td>
</tr>
<tr>
<td></td>
<td>total</td>
<td>10.86</td>
</tr>
<tr>
<td>Total</td>
<td>collaborative</td>
<td>10.40</td>
</tr>
<tr>
<td></td>
<td>individual</td>
<td>11.04</td>
</tr>
<tr>
<td></td>
<td>total</td>
<td>10.67</td>
</tr>
</tbody>
</table>

As Hypothesis 2 refers only to the PF-Coll and PF-Ind conditions an a priori contrast as part of the ANOVA was calculated: In contrast to Hypothesis 2, no significant difference between students trying to collaboratively or individually solve a yet unknown problem prior to receiving explicit instruction was found ($F[1,224]= 1.16$, $p=.28$). Thus, Hypothesis 2 receives no support. Possible reasons will be discussed in the chapter 4.3.

### 4.2.3 Students’ solutions and conceptual knowledge

In order to investigate whether learning mechanism 1, namely students’ prior knowledge activation and differentiation, came to play, the first step was to investigate how many solution ideas students generated across both PF-conditions. Means and standard deviations of students’ solutions are displayed in Table 4.6 (see also Mazziotti, Rummel, & Deiglmayr, 2016). Both PF-Coll and PF-Ind students developed on average 1.5 solution ideas per problem-solving step (solution ideas for problem-solving step 1: $M= 1.71$, $SD=.84$; solution ideas for problem-solving step 2: $M= 1.57$, $SD=.88$; solution ideas for problem-solving step 3: $M= 1.08; SD=.94$). 76 % of the students generated a (single) complete
solution by engaging in all three problem-solving steps (see also chapter 4.1.4 under solution quality).

Table 4.6. Number of student-generated solution ideas and solution quality of the PF-Coll and PF-Ind conditions

<table>
<thead>
<tr>
<th></th>
<th>Number of solution ideas</th>
<th>Solution Quality</th>
<th>Sample size</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><em>M</em></td>
<td><em>SD</em></td>
<td><em>M</em></td>
</tr>
<tr>
<td>PF-Coll</td>
<td>4.53</td>
<td>1.8</td>
<td>4.97</td>
</tr>
<tr>
<td>PF-Ind</td>
<td>4.12</td>
<td>1.92</td>
<td>4.18</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>4.36</td>
<td>1.85</td>
<td>4.64</td>
</tr>
</tbody>
</table>

To further detect *learning mechanism 1* I calculated correlational analyses by using Pearson’s r: I found no significant correlation between the number of students’ solution ideas and students’ conceptual knowledge at posttest ($r(115)=.08, p=.39$). Looking at each PF-condition separately, there were no significant correlations between students’ conceptual knowledge and the number of solution ideas (PF-Coll: $r(66) =.32, p=.79$; PF-Ind: $r(47)=.17, p=.23$).

In order to investigate how students’ collaboration during the initial problem-solving phase of PF affected students’ broad prior knowledge activation and differentiation (cf. learning mechanism 1) an one-factorial ANOVA with the factor condition (PF-Coll vs. PF-Ind) and the number of solution ideas as the dependent variable was conducted. In line with Hypothesis 3, collaboratively learning students generated slightly more solution ideas as compared to individually learning students but not significantly, ($F[1,115] = 1.39, p=.24$). Means and standard deviations of the number of student-generated solutions (and solution quality) are displayed in Table 4.6.

In order to investigate to which extent *learning mechanism 2*, namely students’ awareness of knowledge gaps, evolved, I focused on the quality of students’ initial solution as indicator for the existence of knowledge gaps (serving as a proxy for students becoming

---

2 To keep the process analyses constant across all sources of data (i.e., number of solution ideas, solution quality and students’ collaborative processes) also the analyses of the solution quality and number of solution ideas refers only to those students whose collaborative process data were not lost due to technical issues (as it was the case for six students in the PF-Coll condition). This leads to data of 68 students in the PF-Coll condition rather than to data of 74 students.
aware of these gaps). The solution quality across both PF conditions is with the average number of 4.64 solution quality points slightly over the half of the maximum 9 solution quality points.

To detect how the solution quality is linked to students’ conceptual knowledge at posttest I conducted further correlational analyses and found a low negative and not significant correlation ($r(115) = -.01, p = .90$). Also when looking at each PF-condition separately, I did not find significant correlations between students’ conceptual knowledge and the quality of their solution (PF-Coll: $r(66) = .08, p = .5$; PF-Ind: $r(47) = -.09, p = .56$).

In order to analyze how students’ collaboration during PF’s problem-solving phase affected the extent to which students had knowledge gaps an one-factorial ANOVA with factor condition (PF-Coll vs. PF-Ind) and the solution quality as the dependent variable was conducted (see also Table 4.6). In line with Hypothesis 4, collaboratively learning students descriptively albeit not significantly developed a higher quality solution and might thus have resolved some of their knowledge gaps already during problem solving ($M = 4.97, SD = 2.1; F[1,115]=3.34, p = .07$).

### 4.2.4 Types of collaborative utterances, students’ solutions and conceptual knowledge at posttest

Overall students exchanged over 5074 utterances which could be coded as either interactive, constructive, coordinative, back channeling or off-task. Table 4.7 shows the means and standard deviations of interactive, constructive, active, coordinative, back channeling and off-task utterances. Students produced on average 74.62 total utterances ($SD=31.26$).

**Table 4.7. Means and standard deviations of students’ collaborative utterances in the PF-Coll condition, measured at the individual level (n=68)**

<table>
<thead>
<tr>
<th></th>
<th>$M$</th>
<th>$SD$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interactive</td>
<td>07.50</td>
<td>05.88</td>
</tr>
<tr>
<td>Constructive</td>
<td>14.25</td>
<td>06.53</td>
</tr>
<tr>
<td>Active</td>
<td>03.77</td>
<td>03.32</td>
</tr>
<tr>
<td>Coordinative</td>
<td>23.91</td>
<td>11.44</td>
</tr>
<tr>
<td>Back channeling</td>
<td>12.21</td>
<td>09.15</td>
</tr>
<tr>
<td>Off-Task</td>
<td>12.99</td>
<td>12.09</td>
</tr>
<tr>
<td>Total utterances</td>
<td>74.62</td>
<td>31.26</td>
</tr>
</tbody>
</table>

In order to examine how different types of collaborative utterances are linked to the number of solution ideas as indicator for the supportive role of students’ collaboration for the
evolvement of learning mechanism 1 a number of correlational analyses was conducted (i.e., by using Pearson’s r): I found weak and statistically significant correlations between the number of solution ideas and interactive utterances ($r(66) = .24, p = .048$) and constructive utterances respectively ($r(66) = .34, p = .00$). The more students engaged in interactive and constructive utterances respectively the more solution ideas they generated. There were no significant correlations between the number of solutions and students’ active utterances ($r(66) = .07, p = .59$), back channeling utterances ($r(66) = .17, p = .16$) and off-task related utterances ($r(66) = .07, p = .55$) but with coordinative utterances ($r(66) = .25, p = .04$).

In order to investigate how different types of collaborative utterances are linked to the solution quality as indicator for the existence of knowledge gaps paving the way for learning mechanism 2 further series of correlational analyses (i.e., by using Pearson’s r) were conducted. There were no significant correlations between solution quality and students’ interactive ($r(66) = .05, p = .72$), constructive ($r(66) = -.08, p = .54$) or active utterances ($r(66) = .09, p = .46$) nor coordinative ($r(66) = -.12, p = .33$), back channeling ($r(66) = -.16, p = .18$) and off-task related utterances ($r(66) = -.22, p = .07$).

In order to explore how different types of utterances are linked to students’ conceptual knowledge acquisition I again conducted correlational analyses by using Pearson’s r. I found a weak and statistically not significant positive correlation between the number of interactive utterances a student made during collaboration and student’s conceptual knowledge at posttest ($r(66) = .18, p = .15$). There was the same weak correlational pattern for students’ conceptual knowledge and constructive utterances ($r(66) = .16, p = .19$) and active utterances ($r(66) = .05, p = .71$). In contrast, when investigating how coordinative, back channeling and off-task utterances are connected with students conceptual knowledge small to moderate negative, but statistically also not significant correlations were found (coordinative utterances: $r(66) = -.19, p = .12$; back channeling: $r(66) = -.11, p = .37$; off-task utterances: $r(66) = -.23, p = .06$). Although none of the correlational analyses between different types of collaborative utterances and students’ conceptual knowledge acquisition reached statistical significance the direction of the connection between students’ conceptual knowledge and the different kinds of utterances were in line with the aforementioned literature. Possible reasons for the results will be discussed in the next chapter.
4.3 Discussion

Previous PF-similar studies with young samples (DeCaro & Rittle-Johnson, 2012; Fyfe et al., 2014; Loehr et al., 2014; Matlen & Klahr, 2013) showed mixed results with regard to the effectiveness of problem-solving prior to instruction on students’ conceptual knowledge. While these studies share with the PF approach the delayed timing of instruction, the control condition and the learning outcome measure, these studies differ from PF studies in terms of two design components: An instructor-led comparing and contrasting activity during instruction and students’ collaboration during problem solving. As both design components are expected to trigger learning mechanisms underlying PF the question evolved whether the absence of these design components might explain why learning approaches with delayed instruction after problem solving so far did not consistently show their benefits on young students’ conceptual knowledge acquisition.

Therefore, the goal of study 2 was to adapt the study design of the aforementioned studies in this study (and study 1) by enabling young students to collaborate during problem solving and to receive instruction with instructor-led comparing and contrasting activity in a PF setting. Because the effectiveness of the instructor-led comparing and contrasting activity on students’ conceptual knowledge is well investigated (Loibl & Rummel, 2014a) I again tested for the effects of the (delayed) timing of instruction (including an instructor-led comparing and contrasting activity) on students’ conceptual knowledge. I hypothesized problem-solving prior to instruction (PF) leads to more conceptual knowledge as compared to problem-solving after instruction (DI) (cf. Hypothesis 1). In contrast, so far little is known about the role of collaborative learning for young and older students’ conceptual knowledge within PF. I hypothesized students learning collaboratively during the initial problem-solving phase of PF (followed by instruction) to acquire more conceptual knowledge as compared to individually learning students (cf. Hypothesis 2).

To be able to test these two main hypotheses the larger scale Study 2 systematically varied the factors timing of instruction (problem solving prior to instruction vs. problem solving after instruction) and social form of learning (collaborative learning during problem solving vs. individual learning during problem solving). In order to test Hypothesis 1 I compared both PF-conditions with both DI-conditions. Despite the adaptation of the design of the previous problem-solving prior to instruction studies, the results of Study 2 did not show superiority of PF over DI with young students. Therefore, Hypothesis 1 could not be confirmed. To investigate Hypothesis 2, I compared collaboratively learning students
with the individually learning students in PF. By comparing students from the PF-Coll and PF-Ind conditions I could not find a significant difference. Therefore, Hypothesis 2 receives no support.

In light of Loibl and colleagues’ recent review (2016) a second goal of this study was to systematically analyze learning mechanisms underlying PF. For instance, changes in students’ conceptual knowledge in the end probably only evolve if students sufficiently activated and differentiated their prior knowledge (cf. learning mechanism 1), became aware of gaps in their knowledge (cf. learning mechanism 2) and recognized deep features (cf. learning mechanism 3) during the preceding problem-solving and instruction phase. As students’ collaboration during problem solving and the instruction with comparing and contrasting activity (i.e., two core design components) are expected to trigger the evolvement of these learning mechanisms I conducted a series of processes analyses to detect these mechanisms by making use of multiple sources of data such as the number and quality of student-generated solutions and students’ collaborative utterances. In this context also Hypothesis 3 and 4 were investigated. I will discuss the results of the process analyses for each learning mechanism step-by-step.

### 4.3.1 Students’ prior knowledge activation and differentiation

To explore to which extent student activated and differentiated their prior knowledge (cf. learning mechanism 1) I analyzed the number of solution ideas students generated during the initial problem solving phase of PF and how the number of student-generated solution ideas is linked to students’ conceptual knowledge acquisition (following previous PF studies, e.g., Kapur, 2012, 2014b; Kapur & Bielaczyc, 2012). Across both PF-conditions students generated about 1.5 solution ideas per problem-solving step. In light of the range of possible solution ideas this average number is rather low and thus does not speak for a very broad prior knowledge activation and differentiation. For example, only for problem-solving step 1 (dividing the four pizza Salami equally and fairly amongst six boys) students could have come up with a range of erroneous and correct solution ideas, for example, by making halves (1st solution idea), thirds (2nd solution idea), quarters (3rd solution idea), eighths (4th solution idea), twelves (5th solution idea) and so on. In addition, the number of solution ideas did not significantly correlate with students’ conceptual knowledge acquisition as in the aforementioned PF literature. Therefore, it seems as if learning mechanism 1 did not work appropriately.
However, as there are also other PF studies (Loibl & Rummel, 2014b; Wiedmann et al., 2012) finding no link between the number of student-generated solutions and the learning outcome but replicating the beneficial PF-effect, it seems as the number of student-generated solutions reflects not only students prior knowledge activation and differentiation. Possibly, the number of student-generated solution ideas additionally reflects the extent to which students were persistent to generate solution ideas after repeated failure.

The notion of the number of student-generated solution reflecting students’ persistence to engage in problem solving (Mazziotti, Loibl et al., 2015) might also explain why there was such a low number of solution ideas students generated during the initial problem-solving phase: Possibly, young students as compared to older students may not yet have the same extent of persistence and may not yet have the same repertoire of self-regulated and particularly metacognitive learning strategies at their disposal, such as dealing with failure, control of motivation, control of attention and control of emotions (Kuhl, 1987). Landmann and colleagues (Landmann et al., 2015) even conclude that only at a teenage age (15 to 16 years) students can be expected to learn in a self-regulated manner. In order to investigate this assumption future research should pay more attention to the potential link between students’ self-regulation and the specific requirements of the problem-solving phase in PF.

As collaborative learning during the initial problem-solving phase of PF was meant as the PF component that mainly facilitates learning mechanism 1 a series of additional process analyses investigating the effect of students’ collaboration was conducted: In line with Hypothesis 3, collaboratively learning students descriptively albeit not significantly generated more solution ideas as compared to individual learning students. Diving even deeper into the collaborative process analyses, beneficial collaborative processes such as interactive and constructive utterances significantly correlated with the number of student-generated solution ideas. That means, the more students engaged in interactive utterances and constructive utterances respectively the more solution ideas they generated. In addition, the correlation between interactive as well as constructive collaborative utterances and students’ conceptual knowledge points in the same direction, although it fails to reach significance. Therefore, it seems as students’ collaboration somehow had a positive effect on students’ learning. However, as students in general generated a rather low number of solution ideas there was only a limited number of situations during the problem-solving process in which students’ collaboration might more intensively have helped students to adapt to the typical challenges of the initial problem-solving phase of
This finding might explain why the correlations between interactive and constructive utterances and students’ conceptual knowledge at posttest failed to reach significance and why there was no advantage of collaborative learning over individual learning on the learning outcome (cf. Hypothesis 2).

Interestingly, students also generated more solution ideas the more they coordinated their work \( (r(68)=.25, p=.04) \). Maybe a certain amount of work coordination is required to set the stage for a productive collaborative problem-solving process.

Overall, the results of the series of aforementioned process analyses show a rather low extent to which students’ prior knowledge activation and differentiation evolved (cf. learning mechanism 1).

4.3.2 Students’ awareness of knowledge gaps

By implementing a comparing and contrasting activity during the instruction phase, the design specifically addressed students’ awareness of knowledge gaps (and recognition of deep features of the target domain). For instance, when the instructor compares and contrasts typical student solutions with the canonical solution she draws students’ attention to erroneous student solutions and explains why they are erroneous or incomplete. However, the benefits of the comparing and contrasting activity with regard to learning mechanism 2 might only develop their full potential when students actually developed incomplete solutions by themselves and could thus become aware of their failure already during the initial problem-solving phase of PF. As the solution quality accounts for the existence or absence of gaps in students’ knowledge, I coded and analyzed the quality of students’ initial solution generated during the problem-solving phase and investigated how the solution quality is linked to students’ conceptual knowledge acquisition.

The quality of students’ solution reached approximately half of the maximum of 9 points. Given that students had not received formal instruction yet, this score represents an appropriate quality level: On the one hand, it shows that the problem did not overly challenge the students. On the other hand, this medium quality also shows that students still have gaps in their knowledge and could therefore benefit from the comparing and contrasting activity during the instruction. However, being from this point of view well prepared for the instruction with comparing and contrasting activities did not lead to the
expected superiority of PF over DI (cf. Hypothesis 1). Therefore, I wonder whether students with apparent knowledge gaps actually were able to become aware of these gaps during the initial problem-solving phase of PF.

Possibly, the young students were not yet able to sufficiently evaluate their failed solution ideas as erroneous. In consequence, they were not yet able to become aware of the gaps in their knowledge specifically as young students are expected to have comparably lower evaluation and monitoring capacities as compared to older students (cf. Landmann et al., 2015, when covered under the umbrella of self-regulated learning strategies). This way and in contrast to older students, young students might after failure not have sufficiently reflected about their failed solution idea and thus not analyzed for erroneous features of a given solution idea. Young students might simply have generated another (same quality) solution idea without overcoming some of the erroneous features of the previous solution idea. In other words and following Tawfik and colleagues (2015) failure-based learning cycle the young students might not have adapted their initial mental model in a constructive way.

First hints that older students indeed reflect upon their solution ideas and analyze for erroneous features after evaluated failure comes from a problem-solving process analyses conducted by Kapur (Kapur, 2008). Here students solved ill-structured problems about Newtonian kinematics (followed by well-structured problems). Kapur (2008) identified a likely sequential pattern of different types of process-related activities: After students evaluated their solution they were highly likely to analyze for failure (i.e., conducted a so called problem analysis) and even to criticize the way how they investigated for failure (i.e., engaged in so called problem critique).

However, in contrast to the older students young students might only be able to evaluate their solution ideas when they receive more support. Indeed, in the study conducted by Decaro and Rittle-Johnson (2012) young students solved challenging math equivalence problems and received accuracy feedback after each problem showing them whether their answer was correct or not. Speaking with PF-learning mechanism terms, this feedback helped them to become aware of gaps in their knowledge and to become aware that something was or was not correct with their solution during problem solving. As so far Decaro and Rittle-Johnson’s study (2012) is the only PF-similar study demonstrating the beneficial effect of delaying instruction after problem solving for young students and considering the findings especially of Study 2 future PF research might pay more attention to how
young students evaluated their solutions in the problem-solving phase. More precisely, to investigate the possible interaction between (young) students’ evaluation capacity and the evolvement of students’ awareness of knowledge gaps already during problem solving, future PF research should focus on situations around failure in the problem-solving process, analyze to which extent students actually reflect upon and analyze for failure and how these activities are linked to students’ conceptual knowledge acquisition. For this, a coding scheme would need to be developed.

In order to additionally examine how students’ collaboration during the initial problem-solving phase of PF affected the existence of knowledge gaps as prerequisite for students’ awareness of knowledge gaps further process analyses were conducted. Following Hypothesis 4, I first analyzed to which extent the solution quality differed across the PF-Coll and PF-Ind conditions. On a descriptive (albeit not significant) level, the results show that collaboratively learning students generated a slightly higher quality solution as compared to individually learning students (cf. Hypothesis 4). Thus, collaboratively learning students might indeed have resolved some of their knowledge gaps by discussion. For sure, collaboratively learning students do not have higher evaluation and monitoring capacities as their equally old individual counterparts. However, for the young target group it might have been easier to evaluate some of the learning partner’s solution ideas rather than their own solution ideas (as this would have required to have a certain amount of distance). Nevertheless, considering the small difference between PF-Coll and PF-Ind students with respect to solution quality, the benefits of collaborative evaluation and monitoring might have come only rarely into play.

In a second series of process analyses I investigated how different types of collaborative utterances are linked to solution quality: There were no significant connections between students’ collaborative utterances and solution quality. In line with the ICAP-framework (Chi, 2009; Chi & Wylie, 2014) students active and interactive utterances were positively (albeit not significantly) linked to solution quality. In contrast, and in line with the broader literature about collaborative learning and students’ off-task behavior, students’ off-task, back channeling (i.e. covering no mathematical content) and coordinative utterances were negatively (albeit not significantly) linked to solution quality. Therefore, all connections except for the negative link between students’ constructive utterances and solution quality were in line with the literature. One possible explanation for the counterintuitive link between students’ constructive utterances and solution quality might be that constructive
utterances covered all types of mathematical ideas but did not differentiate between canonical and erroneous ideas whereas the solution quality only accounts for correct solution features.

Against this background I derive the conclusion that students actually had gaps in their knowledge (independent from the social form of learning). Although I currently have no additional data about the extent to which students indeed became aware of gaps in their knowledge, the aforementioned literature and the missing superiority of PF over DI (cf. Hypothesis 1) increase the likelihood of students actually not being aware of their knowledge gaps.

4.3.3 Students’ recognition of deep features of the target domain

By implementing a comparing and contrasting activity during the instruction phase the design additionally addressed students’ recognition of deep features of the target domain (cf. learning mechanism 3). Like the two learning mechanisms just discussed, this third learning mechanism does not seem to have worked appropriately, as the conceptual knowledge posttest measuring for deep features of the target domain fell rather short specifically as the PF students descriptively acquired less conceptual knowledge than the DI students (cf. Hypothesis 1). Recognizing deep features should have been particularly easy for the PF students as they were less novices by the time of the instruction in which the instructor draws students’ attention to deep features as compared to students from the DI condition (Kapur, 2016). Indeed, in the expertise literature (Chi et al., 1981) it is well established that recognizing deep features of the target domain differentiates experts from novices. To be able to more precisely investigate students’ recognition of deep features for future PF studies I propose to implement erroneous worked examples and to let students identify the relevant deep features and the superficial feature of the erroneous worked example.

Due to the interconnectedness of all three learning mechanisms underlying PF in a cascaded way it might overall not be surprising that learning mechanism 3 did not evolve, as already learning mechanism 1 and 2 did not sufficiently show their benefits.

Summing up, it seems as all three learning mechanisms, namely students’ prior knowledge activation and differentiation, students’ awareness of knowledge gaps and recognition of deep features, did not develop to their full potential for the young target group. As the effectiveness of PF is bound to these three core learning mechanisms it
does not seem to be surprising to not find a superiority of PF over DI that has consistently been found for older students (cf. Hypothesis 1).
5 General discussion and conclusion

In the recent years there was a growing body of literature demonstrating the advantage of delaying instruction after problem solving in the form of the Invention approach or the PF approach over so called Direct Instruction approaches (i.e., instruction followed by problem solving) mainly for students’ conceptual knowledge acquisition and transfer performance (Productive Failure: e.g., Kapur, 2010, 2011, 2012, 2014b; Kapur & Bielaczyc, 2011, 2012; Loibl & Rummel, 2014a, 2014b; Invention approach: e.g., Schwartz et al., 2011; Schwartz & Martin, 2004; Belenky & Nokes-Malach, 2012). The beneficial effect of delaying instruction after problem solving is stable across different samples and domains (see aforementioned studies) and is ascribed to three main mechanisms underlying the PF approach and the Invention approach respectively (Loibl, et al., 2016). For instance, in these approaches students are enabled to activate and differentiate their prior knowledge usually through collaborative problem-solving prior to receiving instruction, to become aware of gaps in their knowledge and to recognize deep features of the target domains either through the implementation of contrasting cases during the problem-solving phase of the Invention approach or through the implementation of an instructor-led comparing and contrasting activity during the instruction of the PF-approach.

While the effects of PF (and Invention approaches) are well investigated for middle school, high school or university students, little is known about the effectiveness of PF on younger students’ conceptual knowledge (i.e., around the elementary school age). Comparable studies implementing not PF but similar approaches with delayed instruction and elementary school students have drawn a rather pessimistic picture of the effectiveness of delaying instruction after problem solving for young students (DeCaro & Rittle-Johnson, 2012; Fyfe et al., 2014; Loehr et al., 2014; Matlen & Klahr, 2013). However, the aforementioned comparable studies with younger students differed from PF studies with regard to two core design components: In contrast to studies conducted with older students, in these studies young students so far neither received an instruction with comparing and contrasting activities (first core design difference) nor did they collaborate with a peer during the initial problem-solving phase (second core design difference). The absence of both design components might explain why the beneficial effect of delaying instruction after problem solving so far did not consistently transfer to young students’ conceptual knowledge.
Against this background the overarching goal of this thesis was to shed further light on the effect of the delayed timing of instruction on young students’ conceptual knowledge once the instruction (across both timings) includes the typical comparing and contrasting activity which is known to be superior over instruction without such activity (Loibl & Rummel, 2014a) (first core difference). I hypothesized students solving an equivalent fraction problem prior to receive instruction (i.e., PF) to score higher on the conceptual knowledge posttest as compared to students solving an equivalent fraction problem after receiving instruction (i.e., DI) (Hypothesis 1). With regard to the second core difference, a further goal of this thesis was to systematically examine the effects of students’ collaboration during problem solving for students’ conceptual knowledge acquisition. In contrast to the form of instruction (i.e., with comparing and contrasting activity) the effects of the social form of learning (i.e., collaborative learning vs. individual learning during problem solving) on their conceptual knowledge has not yet been sufficiently touched in the PF literature. I hypothesized students learning collaboratively during the initial problem-solving phase of PF (followed by instruction) to acquire more conceptual knowledge as compared to individually learning students.

In line with Loibl and colleagues (2016) the beneficial effect of PF on students’ conceptual knowledge rests upon the evolvement of the three aforementioned learning mechanisms leading up to the beneficial changes in students’ conceptual knowledge in the end. In consequence, to thoroughly test for the effects of delaying instruction after problem solving on students’ conceptual knowledge with improved study design (i.e., instruction with comparing and contrasting activity and students’ collaboration) thus also required to systematically assess the extent to which students activated and differentiated their prior knowledge (cf. learning mechanisms 1), became aware of gaps in their knowledge (cf. learning mechanism 2) and recognized deep features of the target domain (cf. learning mechanism 3).

To reach these goals, I implemented the problem-solving prior to instruction approach as PF approach enabling students to collaborate during problem-solving and to receive instruction with an instructor-led comparing and contrasting activity. To examine to which extent the improved study design affected students’ conceptual knowledge acquisition and the evolvement of the underlying learning mechanisms two quasi-experimental studies with German fourth and fifths graders were conducted. In both studies the factors
timing of instruction (problem-solving prior to instruction vs. problem solving after instruction) and social form of learning (collaborative learning vs. individual learning) during problem solving were varied.

As for this young age there was no PF learning and test material available yet, one goal of Study 1 was to develop learning and test material about equivalent fractions. Therefore, Study 1 served as feasibility study. It also allowed to develop methods of analysis for detecting the aforementioned learning mechanisms (second goal of Study 1), and, third, to perform an initial test of the two main hypotheses (as the sample size of the feasibility study was low and the variation of both factors was not complete).

By building upon the learning material of Study 1 and by adapting the level of difficulty of the conceptual knowledge posttest and the recruiting strategy (i.e., fifth graders rather than fourth graders) Study 2 then systematically tested for the two main hypotheses and for the evolvement of the underlying learning mechanisms (in this context also for Hypotheses 3 and 4). More precisely, by systematically varying the factors social form of learning (collaborative learning during problem solving vs. individual learning during problem solving) and the factor timing of instruction including an instructor-led comparing and contrasting activity (problem solving prior to instruction, PF vs. problem solving after instruction, DI) I investigated whether delaying instruction after problem solving also works for young students (when the study design is improved as explained above). To analyze the three core learning mechanisms underlying PF a series of processes analyses was conducted in Study 2. In fact, by making use of multiple sources of data such as the number and quality of student-generated solutions and students’ collaborative processes, and by applying the methods being developed with the respective data from Study 1 (cf. second goal of Study 1), I was able to draw a coherent picture about the extent to which students activated and differentiated their prior knowledge (cf. learning mechanism 1), became aware of their knowledge gaps (cf. learning mechanism 2) and recognized deep features of the target domain (cf. learning mechanism 3). In this context, I paid particular attention to the effects of students’ collaboration on both students’ learning process during problem solving and their learning outcome (i.e. conceptual knowledge acquisition) at the end of the problem-solving and instruction phase.

In a first step I will discuss the results across both studies. In doing so, I will first summarize and discuss the results regarding the effects of the delayed instruction including the comparing and contrasting activity on students’ conceptual knowledge and the
evolvement of the learning mechanisms (cf. Hypothesis 1). Then I will discuss the results concerning the effects of students’ collaboration on students’ conceptual knowledge and the evolvement of the underlying learning mechanisms (cf. Hypotheses 2, 3 and 4). In a second step I will point towards limitations and provide an outlook prior to derive pedagogical implications. In a final step I will draw a conclusion.

5.1 Discussion of results

Despite the adaption of the study design, I found mixed results across both my studies. In the initial feasibility study, I found first indications for students learning with PF (i.e., PF-Coll and PF-Ind condition) to acquire significantly more conceptual knowledge as compared to students learning with DI. However, in line with the aforementioned PF-similar studies with a young sample, I did not replicate this finding in the second larger scale study with a complete 2x2 design and much larger sample size. Therefore, the Hypothesis 1 receives only very limited support. Why PF in Study 2 did not evolve its benefit is related to the three core learning mechanisms:

The intensive process analyses revealed that all three learning mechanisms underlying PF, namely students’ prior knowledge activation and differentiation, students’ awareness of knowledge gaps and recognition of deep target features, only poorly came to play. Due to the low number of student-generated solution ideas (1.5 solution idea per problem-solving steps) I conclude students not to have sufficiently activated and differentiated their prior knowledge (cf. learning mechanism 1). Possibly, young students are not as persistent to generate solution ideas after repeated failure as compared to the typical older age group.

With regard to the second learning mechanism, by analyzing the solution quality the process analyses elicited the existence of gaps in student knowledge. However, due to the missing empirical support for an advantage of PF over DI and a potential lower monitoring and evaluation capacity of young students when compared to older students it was interpreted young students not to be sufficiently aware of their own knowledge gaps.

As generally the conceptual knowledge posttest covering the deep target features fell rather short for the PF students also the third learning mechanism evolved only poorly.

The theoretical reason why probably students’ persistence to generate solution ideas after repeated failure and especially students’ evaluation capacity to evaluate a solution idea as failed seem to be so important for the effectiveness for PF becomes clearer when I
decompose the problem-solving process into its single components. For this, I refer to the failure-based cycle developed by Tawfik and colleagues (2015) as well as the three afore-mentioned learning mechanisms (Loibl et al., 2016). Note, against the background of the findings of both studies I conceptualize the three learning mechanisms no longer in a cascade way but in a cyclic fashion.

In the initial problem-solving phase of PF students develop an initial mental model about how to solve the problem at hand by activating and differentiating their prior knowledge (cf. learning mechanism 1). As students fail to come up with the canonical solution they usually experience failure and thus their existing mental model is being challenged (cf. learning mechanism 2). At this point in the failure-based learning cycle (cf. Tawfik et al., 2015) it is of great importance that students recognize their failed solution idea as failed (and thus experience failure more intensively) and to persist to continue on solving the problem. Indeed, by becoming aware of their failure they develop a global awareness of knowledge gaps (cf. learning mechanism 2). This global awareness of knowledge gaps in turn paves the way for students to engage in the next very crucial point in problem-solving (following Tawfik and colleagues, 2015), the analyses for potential reasons for failure. In sense of the Knowledge-Learning-Instruction-framework (KLI-framework; Koedinger et al., 2012) analyzing for failure resembles an understanding and sense-making activity. For this, it is probably the point (i.e., nodal point) in the problem-solving process of PF that particularly facilitates students’ conceptual knowledge acquisition. More precisely, by identifying potential reasons for failure students specify their global awareness of knowledge gaps and thus start to become aware of more specific gaps in their knowledge particularly when they try to differentiate between deep relevant and irrelevant features of the (erroneous) solution at hand (cf. learning mechanism 3). Possibly, by identifying more specific knowledge gaps students have formulated more specific sub-goals they aim to reach. Identifying sub-goals in turn is an important step for solving problems (e.g., Catrambone, 1996) as it probably helps students to determine and specify the cue that initiates the solution search in long-term memory (Raaijmakers & Shiffrin, 1981; Stroebe & Nijstad, 2004) and in the problem-solving space (Newell & Simon, 1972). For this, they might again rely on their prior knowledge (cf. learning mechanism 1) and adapt their (initial) mental model. Based on their adapted mental model they generate a new solution idea and evaluate its viability (again by applying their prior knowledge, cf. learning mechanism 1). As the new solution idea probably is still not the canonical solution students will re-run the cycle again (probably until the problem-solving phase is over).
More theoretical support for the crucial role of analyzing for failure comes from the broader literature about learning with failure such as Piaget’s cognitive disequilibrium (Piaget & Brown, 1985) and Van Lehn’s impasse-driven learning (VanLehn, 1988). Following Tawfik and colleagues summary “the inquiry process is a critical component of any ill-structured problem-solving because it engages the learner to ask additional questions; ascertain potential causes for breakdown; hypothesize the reasons for why they happened; and find evidence to justify the reasons (de Jong & Lazonder, 2014; Herrington, Reeves, & Oliver, 2014; Lazonder, 2014)” (Tawfik et al., 2015, p. 977). Therefore, future research should pay more attention to investigate the role of students’ persistence and evaluation capacity when young students fail to successfully solve problems prior to receive instruction.

With regard to Hypothesis 2, I investigated the effects of another core design component of PF, i.e. students’ peer collaboration during the problem-solving phase (Kapur & Bielaczyc, 2012) which was not incorporated in previous problem solving prior to instruction studies with a young sample (DeCaro & Rittle-Johnson, 2012; Fyfe et al., 2014; Loehr et al., 2014; Matlen & Klahr, 2013). In contrast to Hypothesis 2, collaboratively learning students did not significantly outperform the individually learning students in terms of their conceptual knowledge at posttest across both studies. While in the initial feasibility study there were first indications for collaboratively learning students acquiring slightly albeit not significantly more conceptual knowledge as compared to their individual counterparts, collaboratively learning students from Study 2 acquired less conceptual knowledge as their individual counterparts. Therefore, Hypothesis 2 cannot be confirmed.

In addition, across both studies I found students to make only limited use of the implemented role script as one part of the collaborative support (while the implemented group goal forms the other part) as students only acted in their roles at the very beginning of the problem-solving phase. For this, in the limited amount of in vivo study time possibly the role script only had a sensitizing for how to productively collaborate (if at all) rather than a direct effect (despite the changed recruiting strategy in Study 2). In consequence, one might conclude young students not to have (constantly) collaborated in a fruitful way which in turn might in light of the interaction paradigm (Dillenbourg et al., 1996) be a crucial explanation for the missing superiority of collaborative over individual learning for students’ conceptual knowledge. However, from the intense collaborative process analyses in Study 2 as well as from Angersbach’s work (Angersbach, 2016) I know young
students to engage in an approximately equal amount of fruitful collaborative interaction processes such as interactive and constructive processes as their 14-15 year old counterparts who tried to solve a problem (about variability) prior to receive instruction. Indeed, while young students engaged on average in 29,14 % of total utterance in interactive and constructive utterances\(^3\), the older students engaged on average in 29,54 % of total utterances in interactive and constructive utterances\(^4\). In conclusion, despite young students’ lower experience with collaborative learning and its practices (as I assumed in the discussion of my first study) young students’ collaboration was as good or bad as the collaboration of the typical older age group. Therefore, young students’ ability to (fruitfully) collaborate and to make use of the role script don’t seem to be the crucial explanation of the missing superiority of collaborative over individual learning.

A more likely explanation for the missing superiority of collaborative over individual learning is linked to the general low number of student-generated solution ideas (i.e., 1.5 solution ideas per problem-solving steps in Study 2). Due to this low number of solution ideas the points in the problem-solving process in which students’ collaborative learning might have actually trumped individual learning by interactively analyzing for failure and discussing strengths and weak points of a solution idea were rather low. The results of the series of process analyses hint at potential advantages of students’ collaborative learning at this crucial point in the problem-solving process. For instance, collaboratively learning students generated more solution ideas and also a solution with higher quality as compared to their individual counterparts (albeit not significantly, cf. Hypothesis 3 and 4). And further: The more students engaged in fruitful collaborative processes such as constructive and interactive utterances (by building upon the learning partner’s previous contributions) the more solution ideas they generated (significantly) and the higher was their conceptual knowledge acquisition (albeit not significantly). The number of positive results in favor of students’ collaboration indicates that students somehow benefitted from learning collaboratively. Future PF research should thus continue to investigate the role of students’ collaboration for the effectiveness of PF. More details on this issue are provided in chapter 5.2.2.

\(^3\) A young student on average made 21,75 interactive and constructive utterances and in total he or she made on average 74,65.

\(^4\) An older student on average made 24,69 interactive and constructive utterances and in total he or she made on average 83,56.
5.2 Limitations and outlook

Against the background of the two studies, I would like to point towards some limitations and provide an outlook. In doing so, I will start to discuss issues every researcher is facing when analyzing collaborative learning data and will propose some future lines of collaborative research in PF. In addition, I will provide an outlook about additional factors probably underlying the effectiveness of PF.

5.2.1 Analyzing collaborative learning

In line with the majority of PF studies conducted with older students (e.g., Kapur, 2011, 2012; Kapur & Bielaczyc, 2012; Loibl & Rummel, 2014a, 2014b) also I let young students solve an equivalent fraction problem in a collaborative setting. To better understand the role of this core design difference (i.e., students’ collaboration) for the effectiveness of PF it was analyzed how students’ collaboration affected both the learning outcome (i.e., conceptual knowledge acquisition) and the learning process (cf. process analyses for detecting underlying learning mechanisms). In doing so, I as a number of researchers before me had to deal with a series of typical and not yet overcome challenges regarding the analysis of the complex nature of collaborative learning. Therefore, the current chapter sheds more light on one of these core challenges and explains how the challenge was addressed from a methodological point of view. Thus the goal is to draw a more coherent picture of the results of the doctoral studies.

The core challenge refers to the multilevel and thus non-independent structure of the collaborative data (cf. Janssen, Erkens, Kirschner, & Kanselaar, 2011) as individuals (individual level 1) worked in dyads (group level 2). As students worked in dyads (i.e., PF-Coll and DI-Coll condition) it is likely that individuals who worked together in a dyad are more similar to each other (e.g., with regard to their conceptual knowledge) as compared to individuals working in another dyad. For this, the observations (i.e., individual cases) are highly likely not non-independent from each other (i.e., depend upon each other) (cf. Janssen et al., 2011; Kenny, Kashy, & Cook, 2006). As this challenge for the different kinds of analyses (i.e., hypotheses testing and process analyses) was addressed in different ways I will explain how I dealt with the multi-level and non-independent data structure step-by-step.

For testing the two main hypotheses (investigating the effect of the timing of the redesigned instruction and students collaboration on students’ conceptual knowledge) this
core challenge was ignored as ANOVAs were calculated and work with the basic assumptions of independence of observations and one-level data structure (e.g., Field, 2015). Note that the core challenge concerning multi levels does not only refer to dyads as units of analysis on the upper level but additionally transfers to classes as upper level. Due to the violation of these assumptions, I cannot exclude to have biased estimates. Therefore, one might conclude other statistical models such as multilevel models and hierarchical linear models (HLM, in the following I will use both terms interchangeably) (Bryk & Raudenbush, 1992) which specifically address the multi-level and non-independent data structure to be more appropriate.

The reasons why calculating an ANOVA for testing my hypotheses (especially the first hypothesis comparing the PF with DI conditions) was still a better option than calculating HLM is related to three reasons. First, in terms of the class being the upper level the number of nine classes in Study 2 was not a sufficient requirement to reliably calculate HLMs as some claim to need at least thirty units of analysis on the group level (Kreft & deLeeuw, 1998) or to need a minimum of at least ten units of analysis on both levels (Nezlek, Schröder-Abé, & Schütz, 2006). Second, with view on the dyad forming the upper level across all four experimental conditions (in Study 2) there were not two levels of analysis, as the individually learning students in the PF-Ind and DI-Ind conditions of Study 2 did not learn in small groups at any point in time. Methodologically speaking, there are systematically missing values on the group level. To the best of my knowledge there is no literature yet addressing systematically missing values for HLMs. The third reason for not using HLM for testing the two main hypotheses (with dyads on the upper level) refers to the specific number of members within a group as to some extent the number of two group members within a dyad interferes with the basic rationale underlying HLMs. Usually HLMs estimate regression quotations for each group and an overall regression quotation across all groups. However, as for estimating a group regression quotation for a single dyad there are just two data points at disposal (i.e., conceptual knowledge posttest of dyad member A and B) the predicted/estimated regression equals the empirical one. In other words, all data points are on the regression line and there is no variance that needs to be explained by group membership or any other independent variable. Against this background it does not even make sense to calculate HLMs for the two collaborative conditions (i.e., PF-Coll and DI-Coll conditions). In order to be able to better investigate the impact of multi-levels and non-independent data future PF study might
think about increasing the group size up to 4-5 students. However, only if this is still of theoretical value as theory trumps statistics.

In context of the process analyses investigating student-generated solutions in the PF-Coll condition (i.e., number of solution idea and solution quality) the aforementioned core challenge (i.e., with dyads as the upper level) was addressed in a different way as one of two commonly used strategies was applied. These strategies are the aggregating strategy (making a variable on the individual level to a group variable) and the disaggregating strategy (making a variable on the group level to an individual level variable) (Janssen et al., 2011). For the student-generated solutions the disaggregating strategy was used as the solutions were measured on the group level and treated as variables on the individual level. Methodologically, the disaggregating strategy leads to a higher probability of committing a type I error (cf. Janssen et al., 2011) which cannot be excluded completely when for example the link between the number of solution ideas and students’ conceptual knowledge was investigated. However, as one of the main goals of my studies was to enable students to collaborate in such a way that they don’t have the chance to divide labor I used just one solution sheet per dyad on which the students documented their solution ideas (leading to the number of solution ideas being on the group level).

When analyzing different kinds of students’ collaborative utterances and their link to students’ conceptual knowledge to some extent I overcame the multi-level data structure as I kept both students’ utterances and students’ conceptual knowledge on the individual level. However, by keeping every variable (i.e., students’ collaborative utterances and conceptual knowledge) on the individual level I statistically ignored the impact of the dyad membership (cf. non-independence of observations). In order to address the dyad membership and potential dependencies between for example the frequency of learning partner A’s and B’s interactive utterances and their conceptual knowledge in the future I would like to focus on a statistical model developed by Kenny and colleagues. The so called Actor-Partner-Interdependence-Model (APIM, e.g., Kenny, 1996, Kenny et al., 2006) has frequently been used in the context of social psychology and specifically couple research (e.g., Assad, Donnellan, & Conger, 2007; Barelds & Dijkstra, 2009). It allows to consider the dyadic and thus non-independent relationship of the learning partners (while neither aggregating nor disaggregating the variables at hand nor ignoring the non-independence of observations) by predicting a learning partners’ conceptual knowledge through his own as well as his dyad partner’s collaborative processes. In addition, the APIM opens the opportunity to more precisely analyze whether for example the actor’s
own collaborative interaction processes such as constructively explaining a solution idea have a bigger (or smaller or equally sized) impact on his conceptual knowledge acquisition as compared to his partner’s collaborative processes (e.g., his learning partner constructively explains a solution idea). In other words, the APIM allows to shed new light on the eternal question whether explaining to oneself or explaining to another person facilitates students’ learning outcomes more (here, students’ conceptual knowledge acquisition) (Ploetzner, Dillenbourg, Praier, & Traum, 1999). While research on modelling examples (e.g., Cox, Mckendree, Tobin, Lee, & Mayes, 1999; McKendree, Stenning, Mayes, Lee, & Cox, 1998) would argue all members of the group to learn equally well when one learning partner shows beneficial collaborative behavior (i.e., constructively explaining a solution idea), the broader literature on tutoring instruction (e.g., Graesser, D’Mello, & Cade, 2011) and collaborative learning (e.g., Webb, 1989) predict a superiority of the person who actually gives the explanation rather than the one who receives the explanation.

Naturally, applying such sophisticated statistical models requires fulfilling certain conditions such as a high number of dyads to have sufficient statistical power (for power calculations see also Ackermann & Kenny, 2016). Therefore, future PF researchers who aim to investigate the role of small group’s collaboration by making use of the APIM need to recruit a sufficient number of dyads going beyond the number of 34 dyads (as in the PF-Coll condition of Study 2).

Nevertheless, while the APIM accounts for the multilevel and non-independent data structure and thus offers a solution for dealing with the aforementioned core challenge of analyzing the complex nature of collaborative learning (processes) it still misses to account for another challenge: The dimension of time or better said the consideration of the sequential order of students’ contributions (as in Study 2 students’ utterances). By conducting dynamic multi-level analyses (Chiu, 2008) and developing a representational tool illustrating time-related dependencies (Hmelo-Silver, Jordan, Liu, & Chernobilsky, 2011) Chiu as well as Hmelo-Silver and colleagues offer two examples for research focusing on this additional challenge.

Overall, future PF research and especially the PF research investigating the role of students’ small group collaboration has a series of methodological challenges to consider. In dependence of the specific research goal and focus some of the challenges can be addressed by the aforementioned methods (albeit not all simultaneously).
5.2.2 Further research on collaborative learning within PF

Both studies at hand deliver first empirical hints that students’ collaboration plays a supportive role for triggering the aforementioned learning mechanisms underlying PF and for facilitating students’ conceptual knowledge acquisition in turn. More research is thus need to unfold the role of collaborative learning within PF. A starting point in this direction could be to again vary the social form of learning (collaborative learning during problem solving vs. individual learning during problem solving) and the timing of instruction with instructor-led comparing and contrasting activity (problem solving prior to instruction vs. problem solving after instruction) and could be to (this time) work with the typical older age group. Ideally, the given problem in the envisioned study is thematically embedded in variability as it was the case in Kapur’s study (Kapur, 2012) as well as Loibl and Rummel’s work (Loibl & Rummel, 2014a). This way, future PF researchers would be able to add to examining how students’ collaboration affects’ students learning process during problem solving and their learning outcome (i.e., students’ conceptual knowledge acquisition).

Another approach to shed further light on the role of collaborative learning for PF would be to investigate and compare the number of solution ideas students generated either collaboratively or individually during problem solving and how these solution ideas are linked to students’ conceptual knowledge (as in Study 2; cf. evolvement of learning mechanism 1). More precisely, as students try to come up with as many solution ideas as possible for a problem they don’t know how to solve yet, to some extent problem solving in PF resembles a brainstorming activity (Mazziotti et al., 2015). The brainstorming literature in turn (Stroebe & Nijstad, 2004) predicts a superiority of individual over collaborative brainstorming (albeit in Study 2 it was descriptively the other way around) as in a collaborative setting the generation of associated solution ideas is blocked by turn takings. That means, a learning partner has to wait until it is his or her turn to propose a solution idea (cf. production blocking; Stroebe & Nijstad, 2004). In addition, from the brainstorming literature we also know that once production blocking is removed from the collaborative setting, often due to mutual stimulations collaboratively learning students generate more solution ideas when compared to individually learning students (Stroebe & Nijstad, 2004). Following this line of argumentation, one way to prevent production blocking is in sense of a think-pair-share approach (Lyman, 1981) to first enable students to solve the yet unknown problem individually (by generating solution ideas) and then enabling them to collaboratively continue to solve the problem by sharing solution ideas. This way, the
train of associated solution ideas is due to the initial individual problem-solving phase (followed by a collaborative problem-solving phase and then instructions) not blocked. Therefore, it seems to be vital to combine the beneficial effects of both social forms of learning. In order to also test for this conclusion future PF research would need to implement a 3x2 experimental design varying the factors social form of learning but this time with a combination of both social form of learning within a single problem-solving phase (i.e., individual and collaborative vs. collaborative vs. individual) and timing of instruction with instructor-led comparing and contrasting activity (i.e., problem solving prior to instruction vs. problem solving after instruction).

When examining the role of collaborative learning for the effectiveness of PF a further study focus could be to analyze the effects different group compositions have on students’ conceptual knowledge. This might be a particularly promising approach as Wiedmann and colleagues (2012) found groups with at least one more knowledgeable member to score higher in the posttest as compared to groups with homogenous knowledge members.

Summing up, more systematic research on the role of collaborative learning for the evolvement of the three learning mechanisms underlying PF and on students’ conceptual knowledge is needed. To do so, one would need to additionally take into account the methodological challenges raised above.

5.2.3 Additional factors underlying PF?

In line with findings from my studies and PF-similar studies with young students I lean towards the conclusion not only design components of PF to be decisive for its effectiveness but additionally student characteristics (which are related to students’ age) such as students’ persistence and evaluation capacities to play an important role especially for triggering the aforementioned learning mechanisms. Further support for this notion comes from the Aptitude-Treatment Interaction research (ATI) (Cronbach & Snow, 1981; Snow, 1989) that investigates the effectiveness of a treatment (here a learning approach) in dependence of characteristics of the learner. In this context, I pose the question whether there are additional student characteristics and further factors decisive for the effectiveness of PF which might explain why so far the beneficial effects of learning with delayed instruction did not transfer to a younger sample. In other words, I would like to open the discussion about further potential boundary conditions of productively learning with PF and highlight the need to address these conditions in future research.
One student characteristic that is related to the age and might in sense of ATI negatively interact with characteristics of the treatment (i.e., learning approach with delayed instruction) might be young students’ comparably lower cognitive capacities. According to Gathercole and colleagues (2004) young students as compared to older students have lower cognitive capacities regarding all three components of the tripartite model of working memory (i.e., central executive, phonological loop and visuospatial sketchpad) (for more details about the tripartite model see Baddeley & Hitch, 1974; Baddeley & Logie, 1999). For sure, students need sufficient cognitive capacities to be able to adapt to the specific requirements especially of the problem-solving phase. In line with Tawfik and colleagues key components of failure-based problem-solving the problem-solving phase requires students to repeatedly develop a mental model, analyze for reasons for failure as the initial mental model is being challenged, generate and implement new solutions and finally develop an adapted mental model (Tawfik et al., 2015). Probably, young students as compared to older students due to their lower cognitive capacities don’t withstand these cognitively highly demanding requirements as well as older students. In addition, there is a growing body of literature highlighting a positive link between working memory capacities and mathematics learning (e.g., Passolunghi, Mammarella, & Altoe, 2008). This link supports the assumption that there might in sense of the ATI research be a negative interaction between characteristics of the student (i.e., low cognitive capacities) and characteristics of the treatment (i.e., mathematical domain). Therefore, a starting point for future PF research with a young sample could be to additionally measure students’ cognitive capacities and investigate the potential link to students’ learning outcome.

Although from the comparative perspective younger and older students have in common to be novices in the domain or theme at hand (only intuitive but no formal knowledge, Kapur, 2016), it might be that due to their lower cognitive capacities young students are more likely to use unfavorable (and for novice learners typical) problem-solving strategies such as trial-and-error or mean-ends analyses (Newell & Simon, 1972). Mean-ends-analyses in turn are known to be very cognitively demanding as they require to keep many different but very specific goals simultaneously in mind (i.e., goal specificity effect, Wirth, Künsting, & Leutner, 2009). First hints that young students who first engage in problem solving and then receive instruction tend to use unfavorable problem-solving strategies can be concluded from Fyfe and colleagues’ work (Fyfe et al., 2014). They analyzed the explanations students gave during problem-solving which to some extent display the ap-
plied problem-solving strategies. In fact, students from the problem-solving prior to instruction condition gave more answer-focused and random explanations as compared to the instruction prior to problem-solving condition. Therefore, future research should pay more attention to students’ problem-solving strategies. A step in this direction could be to conduct intensive process analyses and to investigate whether students who use more elaborate strategies also score higher on the posttest as compared to students who use more answer-focused strategies (within the problem-solving prior to instruction condition).

In addition, often the given problem in the initial problem-solving phase of PF is a mathematical story problem and displays to some extent a characteristic of the treatment (e.g., Kapur, 2012; Loibl & Rummel, 2014a). These type of mathematical problems are usually very challenging for students (e.g., Prediger, 2010, Verschaffel, Greer, & Corte, 2000) as they require students to filter relevant from irrelevant text-based information and to transfer these information into mathematical operations. The high level of challenge might hold particularly true for young students as they are not as experienced in (mathematical) text comprehension as compared to older students. However, mathematical education literature investigating how students transfer text-based information into mathematical operations by so called mathematical modelling emphasize the positive link between students’ text comprehension and their mathematical modelling performance (e.g., Schukajlow, 2013). Against this background and line with the ATI research young students’ lower text comprehension and the typical design of the PF problem might again negatively interact with each other and might thus point to an explanation for the missing success of PF. Future PF research should thus assess students’ mathematical text comprehension and analyze how it is linked to students’ learning outcome.

One student characteristic to be decisive for the success of delaying instruction after problem solving is students’ achievement goal orientation. For example, Decaro, Decaro und Rittle-Johnson (2015) analyzed additional achievement goal orientation measures collected in their previous study (DeCaro & Rittle-Johnson, 2012). So far their 2012er study is the only study (at least to the best of our knowledge) demonstrating a superiority of young students solving problems prior to instructions over young students solving problems after instruction with regard to students’ conceptual knowledge. In a nutshell, they argue problem-solving prior to instruction offers an opportunity to investigate how students in a more general way deal with challenging learning events. In line with Dweck’s
work (e.g., Dweck, 1986; Dweck & Leggett, 1988) learning events can either be interpreted as threat to not perform well (i.e., performance goal orientation) or as an opportunity to seek mastery (i.e., mastery goal orientation). They found students’ achievement goal orientation to be decisive not just for students’ learning outcome but additionally for their choice of problem-solving strategies. Students with a mastery goal orientation scored 8% higher on their conceptual knowledge posttest. In contrast, performance goal orientation had no effect on students’ conceptual knowledge. Interestingly, students with a mastery goal orientation applied more elaborated strategies (and less superficial strategies) with increasing complexity of the problem at hand. Further support for the importance of students’ achievement goal orientation comes also from studies conducted with older students (for more details see with older students: Belenky & Nokes-Malach, 2012). Against this background it seems to be vital to control for students’ achievement goal orientation in future studies with young (and older) students at least if students’ achievement goal orientation is not in the center of research interest.

Beyond factors related to characteristics of the target group and of the PF design I would like to additionally consider the classroom environment in which the respective PF study is embedded. Indeed, the kind of classroom norms and practices which are established in the participating school classes might have an impact on the effectiveness of PF as “the nature of socio-mathematical expectations and norms in the classroom influences the extent to which students actually engage in problem solving” (Kapur & Bielaczyc, 2012, p. 52). In addition, also Oser and colleagues emphasize learning with failure ideally happens in an environment in which a so called “Fehlerkultur” (i.e., failure culture) is already well established and implies a productive and appreciating way of dealing with failure (Oser, Hascher, & Spychiger, 1999). In order to be able to additionally account for the impact of the classroom environment a starting point for future research could be to develop a questionnaire that asks for the way how students and teachers generally and specifically in the PF situation deal with failure.

Furthermore, I would like to point to another factor that has not yet been sufficiently touched within the PF and similar literature. In fact, so far it is not clear which role the domain in which the PF setting is situated plays (cf. Loibl et al., 2016). Many PF studies and similar implementations of problem-solving prior to instruction approaches have been conducted within the domains of mathematics (Belenky & Nokes-Malach, 2012; Kapur, 2012, 2014b; Kapur & Bielaczyc, 2011; Loibl & Rummel, 2014b, 2014a; Roll et al., 2011; Schwartz & Martin, 2004) or science (Kapur, 2008, 2011; Kapur & Bielaczyc,
2012; Schwartz et al., 2011) and are thus situated in well-structured domains (as the studies of the thesis at hand). So far two problem-solving prior to instruction approaches were conducted in the domain of psychology and show mixed results regarding the effectiveness of delaying instruction after problem solving (Glogger-Frey et al., 2015; Schwartz & Bransford, 1998). A third study adds to this inconsistent effectiveness when a problem-solving prior to instruction approach is situated within another domain. For instance, Nachtigall and colleagues (Nachtigall et al., 2017) could not replicate the beneficial effect of PF when high school students tried to solve a yet unknown social and educational methodology problem which differs from the typical more well-structured domains of previous PF implementations. Therefore, future PF research should more systematically address the question of how domain characteristics influence the effectiveness of PF.

In addition, as all aforementioned factors and also the discussed learning mechanisms underlying PF refer to cognitive explanations (c.f. Loibl et al., 2016), future PF research might also broaden its perspective to students’ affective states during problem solving as different affective states have differential effects on students learning outcome. Research on learning and affective states has revealed that for example students’ flow and confusion were positively linked to students’ learning whereas boredom was negatively associated to students’ learning outcome (Craig, Graesser, Sullins, & Gholson, 2004). As engaging in complex learning tasks (such as solving the PF problem) often comes with students’ confusion (D’Mello, Lehman, Pekrun, & Graesser, 2014) a starting point for future PF research could be to identify when and how this affective state occurs and how it is linked to students’ learning outcome. This link could be interesting across student age groups.

Overall, the aforementioned additional factors do not provide the ultimate answer why PF as representative for a learning approach with delayed instruction so far did not sufficiently evolve its benefits for young students around elementary school age. However, these factors point towards concepts the relevant PF literature and broader literature about problem-solving prior to instruction approaches has not yet addressed sufficiently but should address in the future.
5.3 Pedagogical implications

In line with my overall findings, young students in contrast to the series of older students do not equally benefit from learning with PF. I discussed how age-related prerequisites of young students such as young students’ potential lower persistence to generate solution ideas after repeated failure and potential lower capacity to evaluate a solution idea as failed to (probably) play a major role for the effectiveness of PF. In parallel to more intensively investigating additional factors underlying PF and specifically the role of students’ potential lower persistence and evaluation capacities for the effectiveness of PF, I would like to propose a pedagogical approach that aims to make PF productive for young learners by specifically addressing these two characteristics.

As young students obviously need more support when compared to older students I would like to emphasize why it is still worth to make PF productive for young learners particularly as this process comes with high implementation costs. As PF is a learning approach supporting the acquisition of conceptual knowledge without comprising for procedural knowledge (Kapur, 2014b) it paves the way for students’ robust knowledge consisting of both types of knowledge (i.e., conceptual and procedural knowledge). As robust student knowledge in turn retains for a long period of time, easily transfers to new situations and accelerates for future learning (Koedinger et al., 2012) it should be the goal for educators and instructional designers across students’ age. Therefore, the benefits of students’ learning with PF-likely outweigh the costs of actually making it productive for young learners.

As young students do not only seem to need more support but are likely, as well, to need just-in-time and adaptive support for being able to evaluate a failed solution idea as failed and to persistently analyze for failure and generate solution ideas right after failure, a first step would be to transfer the problem-solving phase of PF to a computer-supported setting (see also Mazziotti et al., 2015). Similar as in Holmes and colleagues’ work (Holmes, Day, Park, Bonn, & Roll, 2014) in which an invention activity about physics was transferred to a so called Invention Support Environment, also for the envisioned pedagogical approach at hand I would like to transfer the PF-problem-solving phase to a so called exploratory learning environment (ELE; e.g., Mavrikis, Gutierrez-Santos, Geraniou, & Noss, 2013). In such an ELE students could on their own explore different features of the given problem (or the target domain as here fractions) and learn about their relationships by inspecting and manipulating different representations of the domain (Hoyle, 1993). If students still solved the PF problem described in the learning material section of Study
1 and illustrated in Figure 3.2 they might come up with dividing the four pizzas Salami for the six boys into quarters or eights by making circle representations (cf. problem-solving step 1). As this is not yet the canonical solution for this problem-solving step the intelligent component of the envisioned ELE would based on students log data deliver accuracy feedback as in Decaro and Rittle-Johnson’s as well as Fyfe and colleagues’ work (DeCaro & Rittle-Johnson, 2012; Fyfe et al., 2014). This way, students’ potentially lower evaluation capacity would be addressed. To additionally address students’ persistence to continue engaging in problem-solving young students would also be provided with a subsequent prompt saying something like “keep going, sometimes problem-solving is challenging, but you can do it”. In consequence, any time they fail to come up with the canonical solution students would due to the adaptive prompts become aware of their failure (cf. learning mechanism 2) and would thus engage in analyzing for failure - probably the decisive point in problem solving- and would then probably engage in generating another solution idea.

In accordance with the face-to-face implementation of PF also after the ELE-based problem-solving phase an instruction phase would follow. The instruction phase would also be embedded in a computer-supported setting. This way, the overall design of PF would be more consistent and might thus allow students to more easily make connections across learning phases as compared to a cross-media implementation (i.e., problem-solving within ELE an instruction in a classroom-based implementation). I thus propose to implement an instructional video with the typical instructor-led comparing and contrasting activity similar as in Belenky and Nokes-Malach’s work who used instructional videos after students engaged in invention activities (Belenky & Nokes-Malach, 2012).

In addition, while PF already explicitly addresses students’ conceptual knowledge by enabling students to solve problems and explore the underlying target features, so far the typical PF design does not (very) explicitly facilitate students’ procedural knowledge acquisition. Indeed, the current PF design does for example not yet enable students to solve the canonical problem-solving procedure(s) step by step and to receive immediate error feedback (facilitating procedural knowledge). However, in sense of Rittle-Johnson and colleagues’ iterative model “with increases in one type of knowledge leading to gains in the other type of knowledge, which trigger new increases in the first” (Rittle-Johnson, Siegler, & Alibali, 2001, p. 347) students’ procedural knowledge should be facilitated. Therefore, the envisioned computer-based PF approach for young students should additionally implement a third learning phase after receiving instruction (i.e., practice phase)
that would allow students to practice problem-solving procedures step-by-step and to thus gain procedural knowledge. One way to support students to practice procedures during an additional practice phase would be to enable them to learn with an Intelligent Tutoring System (ITS; e.g., Anderson, Boyle, Corbett, & Lewis). ITSs facilitate students’ step-wise problem solving, offer hints and deliver error feedback (e.g., Aleven, McLaren, & Sewall, 2009). In fact, ITSs have shown their effectiveness in a number of studies especially for this type of knowledge (e.g., Pane, Griffin, McCaffrey, & Karam, 2014; VanLehn, 2011).

In order to support students to make links across the aforementioned different kinds learning environments (ELE and ITS), learning activities (problem solving through exploring and practicing procedures) and knowledge types (conceptual and procedural knowledge) in sense of Reigeluth’s Elaboration theory (e.g., Reigeluth, Merrill, Wilson, & Spiller, 1980) I propose to implement a fourth learning phase. During this fourth learning phase students could engage in a synthesizing and reflection activity. For example, a reflection activity could be to let students pose a similar problem (Silver, 1994). This way, students would need to reflect about the single components of the problem they tried to solve during the initial problem-solving phase, learnt to solve during the instruction phase and practiced the procedure with similar problems during the practice phase.

Summing up, the envisioned pedagogical approach proposed to enable young students to learn in a computer-supported setting across four learning phases, namely a problem-solving phase, an instruction phase, a practice phase and a reflection phase. This way, young students would be able to receive age-sensitive and more importantly adaptive support especially during the problem-solving phase as the intelligent ELE component would deliver accuracy feedback and prompts addressing their lower persistence and evaluation capacity—probably the student characteristics that account for the missing effectiveness of PF. After watching an instructional video students are enabled to further practice the previously learned procedure in an ITS by receiving further adaptive support and by engaging in an reflection activity.

When there is more than one problem or one set of problems to be solved, more than one set of deep target features to be learned and more than a canonical problem-solving procedure to be accomplished then the iTalk2Learn project (italk2learn.eu; Mazziotti et al., 2014; Mazziotti, Holmes et al., 2015) provides more information about how to adaptively
sequence the different kinds of learning activities for facilitating complete robust student knowledge about fractions.

## 5.4 Conclusion

It is never nice if hypotheses especially those of doctoral studies are not confirmed. However, what can we learn from this thesis and what contributions did it still make beyond unconfirmed hypotheses.

On a theoretical level the first contribution this thesis made was to further examine whether delaying instruction after problem solving in form of the PF approach is also effective for young students around the elementary school age. In this regard I investigated whether delaying instruction after problem solving is effective for young students once students are enabled to collaborate during problem-solving and to receive instruction with a comparing and contrasting activity. While the effects of the comparing and contrasting activity on students’ conceptual knowledge acquisition are well investigated (Loibl & Rummel, 2014a), little is known about the effects of students’ collaboration on their conceptual knowledge acquisition. In my quasi-experimental studies I thus varied the factors timing of instruction (problem-solving prior to instruction vs. problem-solving after instruction) with the social form of learning (collaborative learning during problem-solving vs. individual learning during problem-solving).

Both quasi-experimental studies showed there is only limited empirical support for the benefits of both delaying instruction after problem solving and students’ collaboration. Also the intensive process analyses detecting step-by-step the core learning mechanisms underlying PF demonstrated students only poorly activated and differentiated their prior knowledge (cf. learning mechanism 1) and seemed to become only poorly aware of existing gaps in their knowledge (cf. learning mechanism 2) and to recognize deep features of the target domain (cf. learning mechanism 3). The rationale to analyze the effectiveness of PF in dependence of underlying learning mechanisms is to the best of my knowledge quite new in the respective literature but in light of Loibl and colleagues recent review (2016) required to thoroughly investigate the effectiveness of PF and similar approaches. In the discussion possible age-related reasons such as young students’ comparably lower evaluation capacity to evaluate a solution idea as failed and lower persistence to continue to engage in problem solving after a potentially frustrating failure experience were identified to probably explain why young students do not yet benefit from problem solving.
prior to instruction. Future PF research should thus in light of the ATI research go beyond investigating design components being decisive for the effectiveness of PF by paying more attention to students’ prerequisites and how they interact with typical design components of PF.

The second important contribution the thesis at hand made was to systematically examine whether collaborative or individual learning during students’ initial problem solving supported students’ conceptual knowledge acquisition. As to the best of my knowledge there are no published studies neither with the typical older age group nor for this young age group that systematically varied the factors social form of learning and timing of instruction within a single study, my thesis sheds further light on the role of collaborative learning for the effectiveness of PF. While students’ collaboration did not show its benefits on the learning outcome level (i.e., students’ conceptual knowledge), the series of collaborative process analyses combining multiple sources of data and making connections between the learning outcome and learning process levels deliver first empirical hints for the supportive role students’ collaboration plays for the effectiveness of PF. To further unfold the potential benefits of students’ collaboration during the initial problem-solving phase of PF a starting point for future research could be to again vary the factors social form of learning and the timing of instruction with the typical older age group.

Based on the interpretations of the findings the third important contribution my research made was to propose an age-sensitive support for students learning with PF. As my findings support the notion that young students highly likely need more and just-in-time support for addressing their probably lower persistence and evaluation capacity during problem solving, I proposed to transfer the existing PF design to a computer-supported setting (i.e., exploratory learning environment). In addition, I proposed to extend the existing PF design by implementing two additional learning phases after problem solving followed by instruction, namely a practice phase and a reflection phase. By enabling students to practice the canonical procedure step-by-step and to receive feedback (i.e. in an ITS) students’ procedural knowledge is in contrast to the majority of previous PF implementations more specifically addressed during the practice phase (i.e., third phase). In addition, by enabling students to make a link across conceptual and procedural knowledge during the reflection phase students have the chance to acquire robust knowledge about the target domain that consists of both conceptual and procedural knowledge, retains for a long period of time, easily transfers to new situations and accelerates for future learning (Koedinger et al., 2012).
6 References


References


E. Wild & J. Möller (Eds.), Pädagogische Psychologie (pp. 49–70). Berlin, Heidelberg: Springer Berlin Heidelberg.


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


References


References


7 Appendixes

7.1 Appendix A: Additional information

7.1.1 Padberg's equivalent fraction problem

Original task in German:

(1) Gleichwertigkeit von \(\frac{3}{6}\) und \(\frac{4}{8}\) in einem konkreten Pizzakontext:


Translation in English:

(1) Fraction equivalence of \(\frac{3}{6}\) and \(\frac{4}{8}\) in a concrete context of pizza:

Pizzeria Caruso and pizzeria Donato bake equally sized pizzas. Pizzeria Caruso divides the round pizzas into 6 equal pieces. Anja buys 3 pieces of pizza. Pizzeria Donato divides the round pizzas into 8 equally sized pieces. Bernd buys 4 pieces of pizza. Who of them gets a greater amount of pizza?

The to-be-chosen answers (distractors) are: Anja, Bernd, both get the same amount of pizza, I don’t know.

7.1.2 Role cards

Role card of the thinker (translated):

You are the thinker. As a thinker you work on the task. In doing so, you explain your thoughts and your approach to the questioner. In this way the questioner is able to understand what the thinker is doing. You should discuss the solution ideas that you already understood und explain them to your learning partner.

You as a thinker could for example start like this:

I think I can begin like this, because....

I think I have to go on like this, because ....

I calculate it in this way, because....

Please remember to act in your role. Only after switching the cards you can change the roles.

Role card of the questioner (translated):

You are the questioner. As a questioner you ask if you won’t understand something or if you can’t comprehend the individual steps of the thinker’s thoughts. Simultaneously you can help the thinker by asking some hint questions. You as a questioner could use for example the following questions:

Why did you calculate it like this?

Could you please explain what you are doing once again?

Hint questions: Why did you calculate it like this, shouldn’t we do it another way? Could we also do it this way?

Please remember to act in your role. Only after switching the cards you can change the roles.
CLAUDIA MAZZIOTTI

EDUCATION
Since 01/2013 Doctoral candidate of the Institute of Educational Research, Ruhr-Universität Bochum, Germany
Since 01/2013 Member of the Research School and member of the Graduate School of Educational Studies, Ruhr-Universität Bochum, Germany
10/2010-09/2012 Master of Arts in Educational Science at Ruhr-Universität Bochum, Germany
03/2009-07/2009 Studies of Italian language and life-long learning at the Università per gli stranieri Siena, Italy
10/2007-09/2010 Bachelor of Arts in Educational Science and Romance Philology at Ruhr-Universität Bochum, Germany
08/1997-06/2006 Evangelisches Gymnasium, Lippstadt, Germany
Abitur (A-level equivalent)

PROFESSIONAL EXPERIENCE
11/2012-09/2017 Research associate: Educational Psychology Research Group, Institute of Educational Research, Ruhr-Universität Bochum, Germany
11/2012-12/2015 Project associate: iTalk2Learn Project, Facilitating Robust student knowledge about fraction with an adaptive learning platform through combining exploratory and structured learning activities (funded by the European Union)
03/2015-04/2016 Research stay: Research on Learning and Instruction Research Group, Eidgenössische Technische Hochschule (ETH), Zurich, Switzerland (funded by Travel scholarship from RUB's Research School)
04/2010-02/2012 Student assistant: Tutor for Introduction into Italian literature at the Roman Seminar, Ruhr-Universität Bochum, Germany
Eidesstattliche Erklärung

Vor- und Zuname: Claudia Mazziotti
Geburtsdatum: 26.08.1986
Geburtsort: Lippe Stadt

Hiermit erkläre ich an Eides statt,

dass ich die vorgelegte Dissertation selbständig und ohne unzulässige fremde Hilfe angefertigt und verfasst habe, dass alle Hilfsmittel und sonstigen Hilfen angegeben und dass alle Stellen, die ich wörtlich oder dem Sinne nach aus anderen Veröffentlichungen entnommen habe, kenntlich gemacht worden sind;

dass die Dissertation in der vorgelegten oder einer ähnlichen Fassung noch nicht zu einem früheren Zeitpunkt an der Ruhr-Universität Bochum oder einer anderen in- oder ausländischen Hochschule als Dissertation eingereicht worden ist.

[Unterschrift]

ORT, DATUM
Bochum, 19.04.2017