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 aforementioned 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 imple-
ment 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 instruc-
tion 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 mem-
bers.

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 effective-
ness 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 effective-
ness 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 asses 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 (Hoyles, 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).