Please cite as follows:

Asterhan, C. S. C. , Schwarz, B. B. & Cohen-Eliyahu, N. (2014). Outcome Feedback during Collaborative Learning: Contingencies between Feedback and Dyad Composition. *Learning & Instruction, 34* (4), 1-10. http://dx.doi.org/10.1016/j.learninstruc.2014.07.003

Outcome Feedback during Collaborative Learning: Contingencies between Feedback and Dyad Composition

Christa S. C. Asterhan, Baruch B. Schwarz & Noa Cohen-Eliyahu

School of Education

The Hebrew University of Jerusalem

Mt Scopus, Jerusalem 91905, Israel

asterhan@huji.ac.il, baruch.schwarz@mail.huji.ac.il, noacoe@gmail.com

In press for *Learning and Instruction*

July, 2014

Abstract

The role of outcome feedback in collaborative learning settings has received little empirical attention. We examined whether outcome feedback improves learning gains in singleton and dyadic learning conditions, while specifying different dyadic pairing options. In a randomized experiment, 496 ninth-graders solved challenging tasks that required fully developed proportional reasoning to be solved correctly. Based on individual pretest performance, each student was assigned to one of three levels of proportional reasoning competence ($Wrong_1$, $Wrong_2$ and Right) and randomly assigned to either work alone or with a ($Wrong_1$, $Wrong_2$ and Right) peer. Half of the dyads and singletons were given the opportunity to empirically test their solutions and received outcome feedback from an objective testing device. The results indicated that when collaboration is considered as a general condition, learners in dyads and singletons profited equally from outcome feedback. When different dyadic compositions are specified, however, the combination of collaborating with a ”Right” partner *and* receiving outcome feedback proved to be particularly powerful. Outcome feedback did not improve learning in any of the other conditions. Furthermore, and contrary to the “two-wrongs-make-a-right-effect”, interaction between two different “Wrong” students did not yield larger gains than other pairing options. The outcomes are discussed in light of existing theories and research.

Keywords: collaborative learning, feedback, dyad composition, proportional reasoning,

1. Introduction

Research has demonstrated powerful effects of feedback for student achievement in individual learning settings (see Hattie & Timperley, 2007; Kluger & DeNisi, 1996, for meta-analyses and overviews). Outcome feedback provides a judgment about the accuracy of the learner’s response. It is one of the simplest and most common types of feedback in educational settings and, compared to control conditions in which no outcome feedback is provided, it is generally associated with positive outcomes (e.g., Kluger & DeNisi, 1996). In contrast to the vast amount of research on feedback in individual settings, outcome feedback is only rarely considered in the collaborative learning literature. In the present study, we investigate the effects of outcome feedback during dyadic and individual learning activities on students’ learning gains. We first discuss the literature on feedback in collaborative settings, and then introduce why the effects of feedback are expected to be dependent on dyadic composition, that is: how dyads are formed based on initial cognitions and competencies.

1.1 Feedback and collaboration

Much of the research on collaborative learning has been based on the idea that peer interaction can be a powerful means for learning if and when peers engage in collaborative sense-making processes (e.g., Chi, 2009; Chi & Menekse, in press). This is evident, for example, when learners explain their thinking to a peer partner (e.g., Coleman, 1998; Van Boxtel, Van der Linden, & Kanselaar, 2000; Webb, Troper, & Fall, 1995), transact on each other’s ideas (e.g., Teasley, 1995), recognize conflicts between their own understanding and other perspectives (Doise, Mugny, & Perret-Clermont, 1975; Howe, 2009), and try to resolve differences through collaborative reasoning (Asterhan & Schwarz, 2007, 2009; Chan, Burtis, & Bereiter, 1997; Howe, Tolmie, Duchak-Tanner, & Rattay, 2000). However, during these collaborative sense-making activities, the correctness of newly developed understandings and problem-solving strategies is often not objectively tested or evaluated by an expert resource. Participants, then, often have no way of knowing whether their solutions are correct other than to rely on their own and their partner’s capacities. This may partly explain why even though many studies have reported positive effects of collaboration, such effects are frequently small and learning outcomes suboptimal, especially for complex topic domains. Similar to individual learning settings, feedback on outcome correctness could then be expected to augment the benefits of peer dialogue, since it provides important information about the particular knowledge that is collaboratively constructed.

How could feedback about outcome correctness be integrated best during collaborative learning? Teachers may scaffold peer discussion by prompting them to engage in sense-making dialogue (Gillies, 2003; Webb, 2009) and gently steer them in certain directions. However, research has also shown that authority and adult evaluations of topic content may undermine the shared meaning-making process that is at the heart of collaborative learning (Hogan, Nastasi, & Pressley, 2000; Webb, 2009). A mid-way should then be found between no feedback at all and authoritative feedback. Such a midway may be provided by activities that allow children to test the correctness of their solutions autonomously with the help of an objective testing device, such as a calculator, scales or other equipment. Peer dialogue and outcome feedback can be alternated in a dialogue-feedback-dialogue sequence: First, children would be required to formulate conceptual knowledge into testable predictions and come to an agreement about which predictions to test. Then they would test and subject these predictions to empirical evaluation (Howe et al., 2000). In those instances where their predictions are disconfirmed, learners may be confronted with compelling evidence that they should reconsider the ideas and explanations that led them to these predictions, thereby creating conflict even when two learners agree on their predictions. Alternatively, in those cases where their predictions are confirmed, the explanation that led to the prediction would be validated. Subsequent sense-making dialogue is needed to interpret the outcomes, particularly in case of conflict. This combination of collaborative sense-making and outcome feedback is likely to be more powerful than either one alone (Tudge, Winterhoff, & Hogan, 1996), especially on tasks for which sense-making dialogue has been shown to be critical, such as conceptual change in complex science and mathematical domains (Asterhan & Schwarz, 2009; Schwarz, Neuman, & Biezuner, 2000). Some have postulated that groups may also be better able to deal with and make use of negative feedback than individuals (e.g., Tindale, 1989).

Yet what is the empirical evidence on such effects? The research available is sparse and has thus far solely focused on groups that consist of two peers (dyads): Schwarz and Linchevski (2007) have shown that ninth-graders who collaborated in dyads and received outcome feedback from a testing device improved their performance on proportional reasoning tasks, whereas singletons who did not receive feedback did not improve. However, the separate effects of feedback and collaboration could not be examined in this study. The separate and combined effects of feedback and dyadic collaboration have been explored in other research, albeit with mixed results: For example, in a study on learning from worked-out examples in college settings, Krause, Stark and Mandl (2009) reported that outcome feedback equally improved performance fo students that either worked alone or in homogenous dyads. In an earlier study, Ellis, Klahr and Siegler (1993) explored the effects of outcome feedback and dyadic collaboration on fifth-graders' use of mathematical rules for decimal fractions. Their results demonstrated that collaborative conditions resulted in superior learning gains *only* when children had access to outcome feedback. Two studies by Tudge and colleagues (Tudge & Winterhoff, 1993a; Tudge et al., 1996) also focused on elementary school students' mathematical reasoning. In direct contrast to Ellis et al. (1993), they found an advantage for dyadic collaboration over individual conditions when children did *not* receive any outcome feedback (Tudge et al., 1996), and an advantage for individual conditions when feedback was provided (Tudge & Winterhoff, 1993a).

Thus, notwithstanding the theoretical rationale for combining collaboration and outcome feedback from equipment, the empirical research is sparse and the evidence available thus far leads to quite different predictions: Based on Tudge and colleagues' research, students would be expected to profit more from feedback when they work alone rather than with a peer partner, whereas based on the study by Ellis and colleagues, they would benefit particularly from the combination of peer collaboration and outcome feedback. According to Krause and colleagues, collaboration does not add anything to the positive effects of feedback.

At closer inspection, however, it appears that each of these studies considered different types of dyadic compositions: Pairing with a partner of an equal, lesser or higher competency level (Tudge et al., 1996), pairing with similar or dissimilar partners of an equal competency level (Ellis et al., 1993) or no specification of dyadic composition at all (Krause et al., 2009). These differences may, then, be responsible for the disparate findings in the literature to date.

1.2 Effects of dyadic composition in collaborative learning

Dyadic composition is based on the student's initial cognitions or levels of competence on the particular concept or problem-solving strategy under investigation. For example, wrong-right dyad configurations (W-R pairs) are made up of one student who has demonstrated a correct understanding of the topic domain prior to the interaction and another who has demonstrated an incorrect understanding of it. Wrong-wrong dyad configurations (W-W pairs) consist of two students with an incorrect prior understanding. These different dyadic configurations in peer collaboration have been the object of many studies by scholars from both the Vygotskian tradition (e.g., Forman & Cazden, 1985; Tudge et al., 1996; Rogoff, 1998) and the Piagetian tradition (e.g., Ames & Murray, 1982; Doise et al., 1975, Doise & Mugny, 1979; Perret-Clermont, 1980). Both theoretical frameworks predict that learning from interaction is not likely to be superior to that from individual learning settings when partners have the same initial cognition, and that a difference of some kind is needed. However, they predict differently *which* type of pairing is more likely to result in cognitive growth (Tudge & Winterhoff, 1993b).

According to Vygotskian scholars, interaction with a more competent peer should lead to better learning, provided that the superior understanding of the more competent peer is accepted and understood through a process of shared meaning-making (e.g., Azmitia, 1988; Garton & Pratt, 2001; Tudge & Winterhoff, 1993a). Neo-Piagetian scholars, on the other hand, have focused on the interactions between two partners with *different* initial cognitions, which are incorrect (W_x-W_y pairs). Several studies have shown that students benefit more from W_x-W_y pairing than from interaction with an R partner (Ames & Murray, 1982; Doise & Mugny, 1979; Doise et al., 1975; Glachan & Light, 1982; Schwarz et al., 2000). Ames and Murray (1982), furthermore, have demonstrated that growth from exposure to a different perspective in W_x-W_y pairs only occurs when children are given opportunities to interact and talk. Most of these studies were conducted with small children on typical Piagetian conservation tasks (but see Schwarz et al., 2000, for an exception).

Results from past investigations on optimal dyadic composition remain, then, inconclusive. Recognizing that what may be responsible for learning in either type of pairing is the extent to which students engage in productive dialogue and collaboratively attempt to establish a shared meaning (Rogoff, 1998), collaborative learning research has then shifted to investigations of peer-to-peer dialogue and how to support it (e.g., Asterhan & Schwarz, 2007, 2009; Gillies, 2003; Howe, 2009; Webb et al., 2008).

Even if it is the quality of dialogue that counts, however, it is still likely that different pairing configurations affect the likelihood that this type of engagement will actually happen (Clark, d'Angelo, & Menekse, 2009; Jermann & Dillenbourg, 2003). Dyadic composition as a design decision for learning tasks is, then, still a relevant topic for the study of learning through peer interaction.

The disparate results of past research on dyadic composition may also be explained by differences in task design, in particular whether learners had access to outcome feedback. Tudge and colleagues (1996) already noted that whilst the effect of feedback has not often been specifically targeted as an object of study in the collaborative learning literature, in some studies the task itself provides students with outcome feedback. This is the case for the Tower of Hanoi task (e.g., Light & Glachan, 1985), for example, where students know whether they have reached the task goal, or not. On others such as those using Piagetian conservation tasks (e.g., Ames & Murray, 1982), the task itself does not provide any feedback on success.

1.3 The combined effects of outcome feedback and dyadic composition

The combined effects of outcome feedback and dyad composition have been examined in two previous studies (Ellis et al., 1993; Tudge et al., 1996), neither of which found that the effect of outcome feedback was dependent on dyadic composition. We already mentioned that these two studies considered different sets of dyadic compositions, however. Furthermore, R-W pairings were not included in either. Thus, for example, in the Tudge et al. study, a W_x student, who was paired with a “more competent” peer may have collaborated with an R partner or with a W_{x+1} partner (a partner who provided a wrong, yet more sophisticated explanation). Distinguishing between these different types of pairing may prove to be important: In collaborating with an R student, W students will not only be exposed to a more sophisticated solution strategy during the discussion phase, but they will also receive empirical confirmation that this strategy leads to the correct solution. This combination is expected to be quite powerful.

The effect of outcome feedback in W-W pairs, on the other hand is less predictable. In collaborating with a more competent W_{x+1} partner, W_x students may initially be convinced by the more sophisticated solution strategy during the discussion phase. Students in a W_x-W_x pair, on the other hand, would be expected to concur with one another quickly, with little discussion required (e.g., Doise & Mugny, 1979; Schwarz et al., 2000). In both cases, however, outcome feedback would ultimately prove their respective solutions as incorrect. This negative feedback could serve as an impetus for further discussion and a re-consideration of the initial solution strategies and explanations (e.g., Limón, 2001), but the emergence of a new, higher-level strategy would have to be accomplished through collaborative discussion only, without further external support.

1.4 The present research and hypotheses

The main aim of the present study is to systematically examine the effects of outcome feedback and collaborative problem-solving on W students’ individual learning gains, while specifying different competence-based dyadic pairings. We recruited a large sample of Israeli ninth-graders. Based on individual pretests, each student was assigned to one of three levels of competence (W_1, W_2 or R). Each W students were assigned to either work alone or with a (W_1, W_2 or R) peer on a challenging task. The task design was based on tenets of cognitive conflict theory. Half the singletons and dyads were given the opportunity to test their solutions with an objective testing device and receive outcome feedback. The other half were not.

The topic domain chosen for this study was proportional reasoning, which involves understanding multiplicative relationships between rational quantities (A/B = C/D). It requires the consideration of relations between relations: Comparing A/B to C/D means comparing a relation between A and B and a relation between C and D (Piaget & Inhelder, 1975). Levels of competence are easily identifiable and hierarchically structured in this domain, which allows for the creation of different dyadic pairing conditions, including W_x-W_x, W_x-W_y and W-R pairings. Proportional reasoning has been described as a watershed competence, a cornerstone of higher mathematics and the capstone of elementary concepts (Lesh, Post, & Behr, 1988). It is important not only in mathematics but also in other academic domains and in everyday life (Boyer, Levine, & Huttenlocher, 2008). Although proportional reasoning is taught systematically in elementary school, it has been shown recurrently that proportional reasoning remains challenging, even for adolescents and adults (e.g., Tourniaire & Pulos, 1985). Performance difficulties on proportional reasoning problems are due in part to an over-extension of counting routines to judgments of proportionality (e.g., Boyer et al., 2008; Mix, Levine, & Huttenlocher, 1999). Accordingly, 6/10 would be judged as greater than 4/6, because 6 is greater than 4. In the current study, adolescents work on a proportional reasoning activity called the Blocks Task. The mathematical problem-solving tasks were designed such that solutions based on non-proportional strategies were disconfirmed by feedback from an objective device, but predictions based on correct proportional reasoning strategies were confirmed. Outcome feedback was provided with the help of scales, which were operated by the students themselves.

The current design allowed us to examine the separate and combined effects of outcome feedback and dyadic pairing condition (alone, with same W, with different W or with R partner) on cognitive growth in proportional reasoning in a comprehensive, controlled study based on a large sample of W students. Our main research questions and hypotheses were as follows:

(1) Does outcome feedback have positive effects on learning in collaborative settings? Based on the rationale outlined above, it was hypothesized that outcome feedback would have a moderate main effect, such that feedback improves learning gains in both collaborative and individual learning settings (*hypothesis 1*).

(2) Does peer interaction lead to improved learning gains? Based on the rationale outlined in the previous sections, it was expected that the answer to this question would

depend on the specific pairing condition: In accordance with socio-cognitive conflict theory and based on past studies (e.g., Ames & Murray, 1982; Schwarz et al., 2000), we hypothesized that interacting with another W student who proposes a different answer (W_x-W_y dyads) would be more advantageous than solitary conditions and interaction with a W peer of the same competency level (*hypothesis 2*). In accordance with Vygotskian tradition, interaction with an R peer was expected to lead to superior learning gains when compared to interactions between two W_x students of the same competency level and to singleton conditions (*hypothesis 3*). As for comparing R-W_x and W_x-W_y pairings, however, the two traditions yield competing predictions: According to Neo-Piagetian theory, W_x-W_y pairings are expected to be more conducive to learning than R-W_x pairings (*hypothesis 4a),* whereas Vygotskian theory would predict the opposite *(hypothesis 4b)*. Since the literature has provided evidence for each of these propositions, we adopted them as competing hypotheses.

(3) Does outcome feedback have a different effect on learning in different dyadic pairing conditions? Based on the aforementioned rationale, we predicted an interaction effect between outcome feedback and pairing condition (*hypothesis 5*). Inter alia, it was expected that for W students, being paired with an R peer and receiving outcome feedback would be a particular powerful combination and lead to substantive cognitive growth when compared to other conditions. As for comparisons between the other pairing conditions (W-W pairings and singleton), our approach was prospective in light of the disparate findings in the literature (Ellis et al., 1993; Tudge et al., 1996) and the mix of obstacles and affordances described here.

2. Method

2.1 Participants

Nine public junior high schools from the Jerusalem and Tel Aviv metropolitan areas (Israel) agreed to participate in the study. The participating classes were typical ninth-grade, integrated mathematics classes, labeled neither ‘‘honors’’ nor ‘‘remedial’’. The entire ninth grade (age 14-15) population of each school (over 600 students) completed a screening (pretest) questionnaire to assess each student's use of problem-solving strategies on proportional reasoning tasks. More than 104 students were excused from further participation when they did not complete the questionnaire, did not provide explanations for their answers, or based their answers on superficial, visual features of the two target shapes only (see Coding section for further details). The remaining 496 ninth-graders (301 boys, 195 girls) used either additive (W_1, N = 196), proto-proportional (W_2, N = 194) or proportional (R, N = 106) reasoning strategies and participated in the intervention task of the study. Students were randomly assigned to pairing conditions (see Design section below for details). Seventy-two R students served as partners to W_1 and W_2 students, whereas 34 others completed the task individually (not further reported here). Since the learning outcomes of non-proportional students are the focus of interest in this study, the final sample included only the W_1 and W_2 students (N = 390).

2.2 Design

After being classified as additive strategy users (W_1), proto-proportional strategy users (W_2) or proportional strategy users (R) based on pretest performance, individual students were randomly assigned to one of five different pairing options: W_1- W_1, W_1- W_2, W_2- W_2, W_1- R and W_2- R (see Table 1 for a distribution of the participants according to the different pairing possibilities). The two different homogeneous W-W pairing options (W_1- W_1 and W_2- W_2) were then collapsed into one condition (W_x- W_x), as were the two different W_x-R pairings (W_1- R and W_2-R). Each dyad or individual was randomly assigned to the outcome feedback condition (with or without). This then resulted in a total of eight different experimental conditions and a 2 (outcome feedback: with or without) X 4 (pairing condition: alone, W_x-W_x, W_x-W_y, W_x-R) experimental design.

Table 1 about here

2.3 Tools

The task items used for the screening, posttest and interaction phase were all adapted from the Blocks Task, originally developed by Harel, Behr, Post and Lesh (1992). Harel and colleagues originally designed the Blocks Taskfor diagnostic purposes, differentiating between students with different mathematical competencies. In the present, adapted version of the Blocks Task items, students are shown 4 three-dimensional block constructions (blocks A, B, C and D), each made up of a number of bricks. The bricks in A and C are of identical color, as are the bricks in B and D. Students are told that the weight of each brick in A and C is identical, and that the same is true of each brick in B and D. In each trial, students are given information about the relation between the two base block constructions A and B (A is heavier than B, B is heavier than A, or they are of equal weight). They are then asked to determine the relation between the weights of the two target blocks, C and D, choosing one of the following four options: “C is heavier than D”, “D is heavier than C”, “They are of the same weight”, or “Impossible to determine”. They are required to provide a verbal explanation for their choice. Figure 1 shows two different Blocks Task items as a examples.

Insert Figure 1 about here

2.3.1 5BlocksTaskTest. Pen-and-paper test compiled of 5 Blocks Task items of increasing difficulty to assess students' proportional reasoning strategies. Table 2 describes the number of bricks in each block, the given relation of weights between A and B, and the relation of weights to be found between C and D. Items 3 and 4 are displayed in Figure 1 as an illustration.

Insert Table 2 about here

The items are designed such that different reasoning strategies lead to different predictions. For example, in item 3 (first item displayed in Figure 1), students who use *visual reasoning strategies* typically claim that C weighs less than D "because it has less bricks in it." Students using an *additive reasoning strategy* would typically predict that C and D weigh the same, because "since A and B weigh the same; C is 4 bricks more than A and D is 4 bricks more than B, then C and D weigh the same". A *proto-proportional reasoning strategy*, on the other hand, typically leads to the conclusion that C weighs more than D. For example, "A has 26 bricks and weighs the same as B which has 27 bricks, and this means that each brick in A weighs more than each brick in B; C is 4 bricks more than A and D is 4 bricks more than B, then C weighs more than D". Students with a *proportional strategy* would also typically predict that C weighs more than D, but with a slightly different reasoning: "If a represents the weight of bricks in A, and b the weight of bricks in B, $26a = 27b$, then $a = 27b/26$, then $30a = 30*27b/26 = 31.15b > 31b$; therefore C weighs more than D."

For this item (item 3), proto-proportional and proportional strategies lead to the same (correct) answer, but the other two strategies do not. In the case of item 4 (second item displayed in Figure 2), the proto-proportional strategy leads to the conclusion that the answer is undetermined (which is incorrect), whereas the proportional strategy leads to the conclusion that C weighs less than D (the correct answer). The 5BlocksTaskTest was designed to assess student's proportional reasoning strategies at pre- and posttest with the help of five test items. In four of the five items, employing an additive strategy would always lead to the wrong answer. The pretest results were also used to determine a learner's reasoning strategy level to create the different pairing conditions for the intervention phase.

2.3.2 Intervention task. The intervention task was designed as a learning task. Depending on conditions, individuals or dyads solved two additional Blocks Task items (I1 and I2). Table 3 shows the configuration of the items.

Table 3 About Here

The students were provided material models of blocks. Students in feedback conditions were provided scales to check their conclusions. For these two items, adopting additive or proto-proportional reasoning strategies leads to a wrong prediction. A priori, wrongness is not inexorable, but it appeared that it was. For example, for Item I2, a typical additive solution would be to notice that Block C has 14 more bricks than A, and that Block D has 15 more bricks than B, and to conclude that weight (C) < weight (D). Students adopting a proto-proportional strategy wrongly conclude that the relation between the weights of C and D is indeterminate.

2.4 Coding procedures

Students' level of proportional reasoning was assessed with the help of a coding scheme that was slightly adapted from Schwarz and Linchevski (2007). Each written response to a test item (5 on pretest, 5 on posttest, and 2 during intervention task for each participant) was assigned to one of four different and mutually exclusive problem-solving strategy categories: In ascending order of reasoning quality, these concerned a visual reasoning strategy (grade: 1), additive strategy (grade: 2), proto-proportional strategy (grade: 3) and full proportional strategy (grade: 4). Descriptions and examples of each strategy category are presented in Table 4. Assignment of a response to a category was based solely on the reasoning used to explain their choice of one of the four alternative answers, and not on the correctness of their response choice.

Table 4 About Here

Ten percent of the entire data set was coded by two independent raters, blind to condition. Inter-rater reliability was Cohen's $\kappa = .925$. The highest strategy level a student used on the pretest version of the 5BlocksTaskTest formed the basis for assessing a student's initial level of proportional reasoning: W_1 (used additive reasoning strategies on each of the 5 pretest items), W_2 (used a proto-proportional strategy at least once, but never a proportional strategy), R (used a proportional strategy at least once). Students who did not use at least additive strategies on all five pretest items (>104 students) were excused from further participation, since previous research on peer interaction and cognitive development showed that in order to show improvement, a student needs at least to be able to articulate clear arguments (also referred to as 'being strategic', see Glachan & Light, 1982; Miller & Brownell, 1975). Performance on pretests and posttests was calculated by the mean grade for the five tasks on each test.

2.5 Procedure

All data collection and experimental interventions were completed locally in each of the nine participating schools. Students participated in the following sequence of activities:

Stage 1: Assessment and selection. The 5BlocksTaskTest was administered in pen-and-paper format to all students in the ninth grade classes to assess their initial level of proportional reasoning. Trained research assistants read aloud the instructions explaining the task. During each of the 5BlocksTaskTest items, the research assistants physically displayed the four relevant constructions (A, B, C and D) for each task in the front of the classroom. Students were free to come up and inspect the four constructions, without touching them. This stage lasted between 25-40 minutes, depending on the amount of time children in each class needed to finish the five tasks.

Stage 2: Intervention. During regular school hours participating students were called to a separate room, adjacent to the participants' classrooms, either individually or in dyads, according to condition. Trained research assistants informed students that they were going to solve two additional tasks and repeated the Blocks Task instructions. Students were shown the four physical block constructions during each task (A, B, C and D). Students in the dyadic condition were instructed to solve the tasks together. They were apprised furthermore that they did not have to reach a consensus, but that they should share ideas and explanations before writing down a solution on one shared sheet of paper. Students in the hypothesis condition also received the following instructions: "After writing down the solution, you can test whether your solution is right or wrong by placing the two target constructions C and D on a scale. If you were wrong you should re-think your solution [together] and try to explain the outcome you received." The research assistants read aloud the instructions and presented the constructions and scales, but refrained from intervening, except to remind students of the instructions when needed.

Stage 3: Post-test assessment. The 5BlocksTaskTest was administered in pen-and-paper format in each classroom after all participating students had completed the intervention phase. All participating students completed the three stages in less than one month, with a 2-3-week interval between the intervention phase and the post-test.

3. Results

Non-adjusted means and standard deviations of pretest and posttest performance per condition are reported in Table 5.

Table 5 about here

All hypotheses were tested with univariate analyses of variance on individual students' mean gains from pretest to posttest. Analyses were conducted with a mixed model (SAS PROC MIXED) with random effects of dyad within condition and of individual within dyad and condition. Two separate models were constructed to examine the research questions: The first model (reported in 3.1) included all 390 W_1 and W_2 participants and tested the overall effects of outcome feedback and collaboration in general, irrespective of pairing composition. A second model (reported in 3.2) was constructed, also including all 390 W_1 and W_2 participants, to compare the effects of specific pairing conditions (singleton, with a same level W partner, with different level W partner, or with a R partner) and the interaction of pairing condition with outcome feedback.

The distribution of studentized residuals was checked for each model separately. Outliers (studentized residuals < -3 or > 3) were locally trimmed from a data set. When the excess kurtosis of a studentized residuals' distribution was greater than one, a SQRT transformation of the dependent variable was used instead. Reported means throughout the remainder of the results section are the estimates of marginal means from trimmed models, adjusted for random effects and are accompanied with the appropriate standard error (SE). Effect sizes for fixed effects are reported in terms of R^2 for Linear Mixed Models (Edwards, Muller, Wolfinger, Qaqish, & Schabenberger, 2008).

3.1 Overall effects of collaboration and outcome feedback

Table 6 presents the estimates of marginal mean gains for W_x learners, according to feedback condition (with or without) and whether students worked in singleton or collaborative conditions, irrespective of dyadic composition. In accordance with hypothesis 1, a marginally significant effect was found for outcome feedback, F (1, 358) = 3.78, p = .053, with an effect size of R^2 = .04: Students in the outcome feedback condition showed larger gains than those who did not. No main effect of collaborative condition was found, F (1, 358 < 1), *ns*, and the two factors were not found to interact, F (1, 358) < 1, *ns*.

Table 6 about here

3.2 Specific effects of pairing condition and outcome feedback

Next, collaboration as a general condition was further subdivided in the following pairing options: homogeneous W-W pairings (W_x–W_x), heterogeneous W-W pairings (W_x–W_y), and pairing with an R partner (W_x–R). Figure 2 plots the estimated mean gain scores for each of the eight conditions.

Figure 2 about here

When dyadic composition was further specified, no main effect was found for outcome feedback, F (1, 236) = 2.53, p = .113. A main effect for pairing condition was found, F (2, 231) = 12.31, p < .0001, with an effect size of R^2 = .21. Post-hoc analyses (with Tukey-Kramer adjustments) showed that, in conformance with hypothesis 3, being paired with an R student (M = .38, SE = .04) indeed resulted in larger learning gains compared to being paired with a same-level W peer or working individually, t (223) = 6.08, p < .001 and t (280) =2.94, p = .019, respectively. In contradiction to hypothesis 4a, but in confirmation of competing hypothesis 4b, being paired with an R partner also resulted in larger gains than being paired with a different-level W student, t (215) = 3.05, p = .014. In contradiction to hypothesis 2, individuals in W_x-W_y pairing conditions did not show larger learning gains compared to individuals in W_x-W_x pairing or singleton conditions, t (191) = 2.31, p = .010 and t < 1, respectively.

In conformance with hypothesis 5, a significant interaction effect was found between pairing condition and outcome feedback condition, F (2, 231) = 3.26, p = .022, with an effect size of R^2 = .06. Tukey-Kramer tests for multiple comparisons revealed the following pattern: In the feedback condition, students who were paired with an R peer (M = .53, SE = .06) gained significantly more than students who were paired with a same-level W peer (p < .001), paired with a different-level W peer (p < .001) or worked individually (p = .010). There were no significant differences between the latter three conditions. In the no feedback conditions, on the other hand, being paired with an R peer did not have any advantage over either being paired with a same-level W peer (p = .082), with a different level W peer (t < 1), or working individually (t < 1). None of the other comparisons between pairing conditions within either of the two feedback conditions yielded significant differences.

Comparisons between feedback and no-feedback conditions within a given pairing condition showed the following pattern: W learners who were paired with an R student gained more when they obtained feedback from the scales about the correctness of their prediction, t (232) = 3.40, p = .018. Comparisons between feedback and no-feedback in the other three pairing conditions, on the other hand, did not reveal any differences in learning gains.

Taken together, it seems that for W_x learners as a group neither outcome feedback alone nor the pairing with an R peer alone resulted in learning gains, but rather only the combination of the two. This is further supported by the finding from post-hoc comparisons that W_x learners in the outcome feedback with an R partner showed statistically significant higher gains scores than students in each of the other 7 conditions. None of the other comparisons reached significance.

4. Discussion

When collaboration was considered as a general condition, without specification of dyadic composition, receiving outcome feedback equally improved students' learning gains in singleton and collaborative conditions (Hypothesis 1). Collaboration in general, without further specification of different dyad composition types, was not found to result in superior learning outcomes, when compared to individual conditions.

When collaboration as a general condition was subdivided into different dyadic compositions (W_x-W_x, W_x-W_y, and W_x-R), however, a main effect of pairing condition on learning gains was found, with W_x students who interacted with R partners outperforming learners in W_x-W_x, W_x-W_y and singleton condition (Hypotheses 4b and 3). Contrary to expectations, interaction in heterogeneous pairs (W_x-W_y) did not lead to higher learning gains than interaction in homogeneous pairs (W_x-W_x) (Hypothesis 2). Finally, the main effect of receiving outcome feedback disappeared when different dyadic compositions were specified. Instead, the effect of outcome feedback interacted with pairing condition (Hypothesis 5): Outcome feedback only benefited those W students who were paired with an R student. It did not have a significant effect on learning in any of the other pairing conditions (W_x -W_x, W_x-W_y, and singleton).

4.1 Integration with past research on outcome feedback in collaborative learning

While there is abundant research on the advantages of outcome feedback in individual learning settings, a few studies have empirically explored its role in collaborative learning settings and with inconsistent results (e.g., Ellis et al., 1993; Krause et al., 2009; Schwarz & Linchevski, 2007; Tudge et al., 1996). Realizing that this inconsistency in outcomes could potentially be attributed to the fact that different studies considered different types of dyadic compositions in their respective collaborative conditions, the present study included W_x- W_x, W_x-W_y, and W_x–R pairs as well as singleton conditions. The finding that outcome feedback equally benefited students in the singleton and in the collaborative condition when the latter did not further specify dyadic composition type, confirms results reported by Krause and colleagues (2009). However, the finding that the effect of outcome feedback was found to interact with dyadic composition emphasizes the importance of specifying differences in group composition, when examining the role of outcome feedback in collaborative learning.

The finding that W students only profited from outcome feedback when they interacted with an R student, but in *none* of the other pairing conditions is somewhat surprising, especially in light of the overall positive effects of outcome feedback in individual conditions (Hattie & Timperley, 2007; Kluger & DeNisi, 1996). Following instructional design theories emphasizing the role of cognitive conflict in learning (see Limón, 2001) one may have expected that negative feedback triggers discussion and cognitive re-structuring, even when two learners initially agreed or shared the same misconceptions. The present findings seem to suggest that outcome feedback may have played quite a different role:

Instead of being a trigger for cognitive restructuring, it provided a means to objectively validate a more sophisticated solution strategy proposed by a more competent peer during the interaction. Without this feedback, the more sophisticated solution strategy offered by the R peer may have remained no more than an alternative solution to a problem.

The current study does not provide data on the students' subjective experiences during the learning task. We can, therefore, only speculate about the reasons why outcome feedback did not improve learning gains for W-W dyads. First of all, it is possible that students simply did not accept that their solution strategy was wrong, in spite of the negative feedback. Chinn and Brewer (1993) have documented multiple ways in which learners explain disconfirming information without accepting that they were wrong. Second, it is possible that instead of eliciting collaborative sense-making discussion, the negative feedback affected students' motivation to continue, resulting in abstention from further engagement.

Third, it is also possible that W students who participated in singleton or W-W pairing conditions did in fact experience cognitive conflict when their predictions were were disconfirmed, but that they lacked the tools to resolve it correctly through dialogue or private deliberation only. Previous research has shown that the experience of conflict during collaborative learning interactions may not necessarily lead to immediate learning gains, and yet may prime individual learners to be more attuned to and to gain from subsequent learning opportunities (Howe, McWilliam, & Cross, 2005). Research on phenomena such as productive failure (Kapur, 2011, 2012) and learning through invention (Schwartz & Martin, 2004) furthermore shows that the combination of autonomous, collaborative attempts to discover mathematical procedures for solving complex problems can have powerful effects, if they are followed by detailed explanations of the correct solution. Two recent studies on affective states and learning (D'Mello, Lehman, Pekrun, & Graesser, 2014) have shown that confusion can be a catalyst for learning, but only when learners are given adequate scaffolds to resolve it. Extrapolating from these studies to the present work, we may arrive at the following tentative conclusion: Even if the negative feedback in collaborative conditions induced cognitive conflict or the awareness that one's current solution strategies are inadequate, without further follow-up scaffolding or instruction, W students are not likely to arrive at the correct proportional reasoning strategy all by themselves, even when they are working in heterogeneous pairs. Future research should focus on uncovering the potentially different roles of disconfirming feedback in these dyadic compositions, for example, by collecting detailed process data of the interaction phases.

4.2 Integrating with past research on dyadic composition in collaborative learning

The analyses presented in this study reveal that collaboration with an R student was more beneficial than working alone or collaborating in W-W dyads. At first impression, this finding seems to corroborate Vygotsky-inspired theories of collaborative learning according to which interaction with a more competent partner is more effective, since learners are exposed to higher-order reasoning strategies (e.g., Garton & Pratt, 2001; Rogoff, 1998; Tudge et al., 1996). Further analyses revealed, however, that interaction with an R partner exclusively improves learning when the outcome feedback confirms the solution proposed by that partner. When no feedback was available, students paired with an R peer showed similar gains to those who worked alone or were paired with a different-level or same-level W peer. The pattern that emerges from these findings, then, seems to underline the importance of the combination of exposure to higher-order reasoning strategies *and* the validation of the correctness of these strategies by an objective test. This is not an additive effect, since neither the discussion of higher-order reasoning strategies (pairing with R peer, no feedback) nor the conflict created by the disconfirmation of incorrect predictions (pairing with another W peer, with feedback) alone led to substantive learning gains.

The finding that, overall, W_x-W_y pairing conditions did not lead to an advantage over others corroborates in part findings reported by others (Ellis et al., 1993; Russell, Mills, & Reiff-Musgrove, 1990). It stands in apparent direct contradiction, however, to neo-Piagetian accounts of collaborative learning processes (e.g., Doise & Mugny, 1979) and research reporting on the “two-wrongs-make-a-right” phenomenon (e.g., Ames & Murray, 1982; Glachan & Light, 1982; Schwarz et al., 2000). We would like to offer the following explanations: Proportional reasoning is a domain in which the different reasoning strategies are hierarchically structured. Accordingly, W_x-W_y pairings were composed of two students with different solution strategies that were both wrong, but one was more sophisticated than the other. In other domains, such as decimal fractions (Ellis et al., 1993; Resnick et al., 1989; Schwarz et al, 2000), evolution (Asterhan & Schwarz, 2007) and conservation strategies (Ames & Murray, 1982), to name a few, incorrect strategies or concepts may be qualitatively different yet equally wrong. The empirical studies that have reported on a “two-wrongs-make-a-right” effect have compared W_x-W_y pairings where x and y are of equally wrong, yet qualitatively different solution strategies. It is not possible, therefore, to directly compare our findings on W_x-W_y pairings to those that reported on the “two–wrongs-make–a-right” effect.

For example, in the Schwarz et al. (2000) study, the W_x-W_y pairs consisted of two children who each had a qualitatively different misconception profile about decimal fractions, such as ‘disregarding zeros in a decimal part’ (e.g., $4.7 < 4.08$) and ‘the number with more decimal digits is larger’ (e.g., .4.8 < 4.68). Thus, within a W_x-W_y pair, W_x may be wrong about his own misconceptions, but can effectively identify and correct the misconceptions of W_y, and vice versa. Learning in these conditions, therefore, may also have been a matter of gaining from each other’s relative expertise in a complementary matter, rather than learning from a conflict between two incommensurate, wrong explanations and having to generate a different, more sophisticated explanation all by themselves.

In summary, based on the current findings and existing research it is imperative that future research efforts specify in detail the partner’s competence level, the target learner’s own competence level, the relation between these two levels of competency (e.g., complementary, hierarchically), the task demands, and the particular feedback that they receive from the testing device (e.g., confirmation, disconfirmation, none). Contingencies between these factors create different learning opportunities for learners of different competence levels. Research efforts should be specific about these contingencies, in order to progress towards an instructional framework that articulates not whether, but rather *when*, *for whom* and *with whom* collaboration and outcome feedback are likely to benefit learning.

4.4 Limitations and future directions for research

In addition to those already mentioned, the following limitations and directions for future research are also identified: First of all, the findings reported here may be limited to the particular group size this study and previous research on outcome feedback and collaboration (Ellis et al., 1993; Krause et al., 2009; Schwarz & Linchevski, 2007; Tudge et al., 1996) has focused on, namely the dyad. In peer groups of larger sizes, other social processes may play a more prominent role and determine outcomes, such as pressure of majority opinions. Secondly, previous studies varied greatly not only in terms of the type of dyadic pairings they considered, but also in terms of age group. The age of participants ranged from first-graders (Ames & Murray, 1982; Russell et al., 1990; Tudge et al., 1996) to fifth-graders (Ellis et al., 1992) to tenth-graders (Schwarz et al., 2000). Given that self-regulatory skills develop considerably in elementary school (e.g., Zimmerman & Martinez-Pons, 1990), it is possible that very young students have difficulty coordinating both collaborative and self-regulatory processes, but that these capabilities improve with age. Future studies should include different age groups to explore such potential age differences.

Thirdly, the choice to focus on proportional reasoning constrained the findings of this study to domains in which differences in competency level are hierarchically organized. Consideration of a particular type of dyadic composition was precluded here, namely, the pairing of two students that use different yet equally wrong strategies. Consequently, the findings of the present study should not only be replicated in a different domain with a similar structure of competency levels (hierarchical), but also be tested in a domain with a different competency level structure.

Finally, an additional aspect of domain knowledge concerns the extent to which (R or W) students are confident about the accuracy and correctness of their own knowledge. The role of such confidence has not often been considered in collaborative learning, but could be potentially important. For example, in previous studies that reported negative outcomes of learning through interaction with R partners (e.g., Ames & Murray, 1982; Doise & Mugny, 1979; Schwarz et al., 2000), the lack of gains was explained by the fact that the latter were highly confident, dominated the interaction and invested little effort in explaining their solutions, to themselves or to their partners. In the current study, on the other hand, students in the R category deployed a full-fledged proportional strategy in at least one of the pretest items. The use of numerical, multiplicative strategies was not needed for *all* individual pre- and posttest items to be solved correctly; some of them were solvable with verbal reasoning only. Finally, the intervention tasks were challenging even for the R students. Thus, although we do not have data on measures of confidence, R students in the present study may have been less confident in general, and therefore more willing to engage in collaborative sense-making dialogue with their partners, than in the aforementioned studies. Future studies should include measures of confidence and detailed process data from the interaction process to enable better comparisons across studies.

5. References

Ames, G. J., & Murray, F. B. (1982). When two wrongs make a right: Promoting cognitive development through cognitive conflict. *Developmental Psychology, 18*(6), 894-897. doi:10.1037/0012-1649.18.6.894.

Asterhan, C. S. C., & Schwarz, B. B. (2007). The effects of monological and dialogical argumentation on concept learning in evolutionary theory. *Journal of Educational Psychology, 99*(3), 626-639. doi:10.1037/0022-0663.99.3.626.

Asterhan, C. S. C., & Schwarz, B. B. (2009). Argumentation and explanation in conceptual change: Indications from protocol analyses of peer-to-peer dialogue. *Cognitive Science, 33*(3), 374-400. doi:10.1111/j.1551-6709.2009.01017.x.

Azmitia, M. (1988). Peer Interaction and problem solving: When are two heads better than one? *Child Development*, *59*(1), 87-96.

Boyer, T. W., Levine, S. C., & Huttenlocher, J. (2008). Development of proportional reasoning: Where young children go wrong. *Developmental Psychology, 44*(5), 1478–1490. doi:10.1037/a0013110.

Chan, C., Burtis, J., & Bereiter, C. (1997). Knowledge building as a mediator of conflict in conceptual change. *Cognition and Instruction, 15,* 1-40. doi: 10.1207/s1532690xci1501_1.

Chi, M. T. H., & Menekse, M. (in press). Dialogue patterns in peer collaboration that promote learning. In: L. B. Resnick, C. S. C. Asterhan, and S. Clarke (Eds), *Socializing Intelligence through academic talk and dialogue*. New York, NY: Routledge, AERA books.

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

Chinn, C. A., & Brewer, W. F. (1993). The role of anomalous data in knowledge acquisition: A theoretical framework and implications for science instruction. *Review of Educational Research, 63*, 1-49. doi: 10.3102/00346543063001001.

Coleman, E. B. (1998). Using explanatory knowledge during problem solving in science. *Journal of the Learning Sciences, 7,* 387–427. doi: 10.1080/10508406.1998.967205.

Clark, D. B., D’Angelo, C. M., & Menekse, M. (2009). Initial structuring of online discussions to improve learning and argumentation: Incorporating students' own explanations as seed comments versus an augmented-preset approach to seeding discussions. *Journal of Science Education and Technology, 18*(4), 321-333. doi: 10.1007/s10956-009-9159-1

D’Mello, S., Lehman, B., Pekrun, R., & Graesser, A. (2014). Confusion can be beneficial for learning. *Learning and Instruction, 29,* 153-170. doi**:**10.1111/1467-8624.00081.

Doise, W., & Mugny, G. (1979). Individual and collective conflicts of centration in cognitive development. *European Journal of Social Psychology, 9*(1), 245–247. doi:10.1002/ejsp.2420090110.

Doise, W., Mugny, G., & Perret-Clermont, A.-N. (1975). Social interaction and the development of logical operations. *European Journal of Social Psychology, 6*(3), 367–383. doi:10.1002/ejsp.2420050309.

Edwards, L. J., Muller, K. E., Wolfinger, R. D., Qaqish, B. F., & Schabenberger, O. (2008). An R^2 statistic for fixed effects in the Linear Mixed Model. *Statistical Medicine, 27*(29), 6137-6157. doi:10.1002/sim.3429.

Ellis, S., Klahr, D., & Siegler R.S. (1993). The effects of feedback and collaboration on changes in children's use of mathematical rules. Paper presented at the Biennial Meetings of the Society for Research in Child Development, New Orleans.

Forman, E. A., & Cazden, C. B. (1985). Exploring Vygotskian perspectives in education: The cognitive value of peer interaction. In J. V. Wertsch (Ed.), *Culture, communication and cognition: Vygotskian perspectives* (pp. 323-347). New York: Cambridge University Press.

Garton, A. F., & Pratt, C. (2001). Peer assistance in children’s problem solving. *British Journal of Developmental Psychology, 19*(2), 307–318. doi:10.1348/026151001166092.

Gillies, R. M. (2003). The behaviors, interactions, and perceptions of junior high school students during small-group learning. *Journal of Educational Psychology, 95*(1), 137-147. doi:10.1037/0022-0663.95.1.137.

Glachan, M., & Light, P. (1982). Peer interaction and learning: Can two wrongs make a right? In G. Butterworth and P. Light (Eds.), *Social cognition: Studies in the development of understanding* (pp. 238-262). Chicago: University of Chicago Press.

Harel, G., Behr, M., Post, T., & Lesh, R. (1992). The Blocks task: Comparative analyses of the task with other proportion tasks and qualitative reasoning skills of seven-grade children in solving tasks. *Cognition and Instruction, 9*(1), 45-96. doi:10.1207/s1532690xci0901_2.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*, 81-112. doi: 10.3102/003465430298487

Hogan, K., Nastasi, B. K., & Pressley, M. (2000). Discourse patterns and collaborative scientific reasoning in peer and teacher-guided discussions. *Cognition and Instruction, 17*, 379-432. doi: 10.1207/S1532690XCI1704_2.

Howe, C. (2009). Collaborative group work in middle childhood: Joint construction, unresolved contradiction and the growth of knowledge. *Human Development, 52*(4), 215-239. doi:10.1159/000215072.

Howe, C., McWilliam, D., & Cross, G. (2005). Chance favors only the prepared mind: Incubation and the delayed effects of peer collaboration. *British Journal of Psychology, 96*, 67-93. doi: 10.1348/000712604X15527.

Howe, C., Tolmie, A., Duchak-Tanner, V., & Rattay, C. (2000). Hypothesis-testing in science: Group consensus and the acquisition of conceptual and procedural knowledge. *Learning and Instruction, 10*(4), 361-391. doi:10.1016/S0959-4752(00)00004-9.

Jermann, P. & Dillenbourg, P. (2003) – Elaborating new arguments through a CSCL scenario. In J. Andriessen, M. Baker, and D. Suthers (Eds.) *Arguing to Learn: Confronting Cognitions in Computer-Supported Collaborative Learning Environments* (pp. 205-226). Kluwer, Amsterdam, NL.

Kapur, M. (2011). A further study of productive failure in mathematical problem solving: Unpacking the design components. *Instructional Science, 39*(4), 561-579. doi:10.1007/s11251-010-914-3

Kapur, M. (2012). Productive failure in learning the concept of variance. *Instructional Science. 40(*4), 651-672. doi: 10.1007/s11251-012-9209-6

Kluger, A. N., & DeNisi, A. (1996). Effects of feedback intervention on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, *119*, 254–284. doi:10.1037/0033-2909.119.

Krause, U. M., Stark, R., & Mandl, H. (2009). The effects of cooperative learning and feedback on e-learning in statistics. *Learning and Instruction, 19*, 158-170. doi: 10.1016/j.learninstruc.2008.03.003.

Lesh, R., Post, T., & Behr, M. (1988). Proportional reasoning. In J. Hiebert & M. Behr (Eds.), *Number Concepts and Operations in the Middle Grades* (pp. 93-118). Reston, VA: Lawrence Erlbaum & National Council of Teachers of Mathematics.

Light, P., & Glachan, M. (1985). Facilitation of individual problem solving through peer interaction. *Educational Psychology, 5*, 217-225. doi: 10.1080/01443418500503.

Limón, M. (2001). On the cognitive conflict as an instructional strategy for conceptual change: A critical appraisal. *Learning and Instruction, 11*, 357-380. doi: 10.1016/S0959-4752(00)00037-2.

Miller, S., & Brownell, C. (1975). Peers, persuasion and Piaget: Dyadic interaction between conservers and non-conservers. *Child Development, 46*(4), 992–997.

Mix, K. S., Levine, S. C., & Huttenlocher, J. (1999). Early fraction calculation ability. *Developmental Psychology, 35*(5), 164-174. doi: 10.1037/0012-1649.35.1.164.

Perret-Clermont, A.-N. (1980). *Social interaction and cognitive development in children*. London: Academic Press.

Piaget, J., & Inhelder, B. (1975). *The origins of the idea of chance in children*. New York: Norton.

Resnick, L. B., Nesher, P., Leonard, F., Magone, M., Omanson, S., & Peled, I. (1989). Conceptual bases of arithmetic errors: The case of decimal fractions. *Journal for Research in Mathematics Education, 20*(1), 8–27. doi**:** 10.2307/749095.

Rogoff, B. (1998). Cognition as a collaborative process. In W. Damon and D. Kuhn (Eds.), *Handbook of child psychology,* Vol. 4, 5th Ed. (pp. 679 – 744). New York: Wiley.

Russell, J., Mills, I., & Reiff-Musgrove, P. (1990). The role of symmetrical and asymmetrical social conflict in cognitive change. *Journal of Experimental Child Psychology, 49,* 58-78.

Schwartz, D. L., & Martin, T. (2004). Inventing to prepare for future learning: The hidden efficiency of encouraging original student production in statistics instruction. *Cognition and Instruction*, *22*(2), 129-184. doi: 10.1207/s1532690xci2202_1.

Schwarz, B. B., & Linchevski, L. (2007). The role of task design and of argumentation in cognitive development during peer interaction: The case of proportional reasoning. *Learning and Instruction*, *17*(5), 510-531. doi:10.1016/j.learninstruc.2007.09.009.

Schwarz, B. B., Neuman, Y., & Biezuner, S. (2000). Two wrongs may make a right...if they argue together! *Cognition and Instruction, 18*(4), 461-494. doi: 10.1207/S1532690XCI1804_2.

Teasley, T. D. (1995). The role of talk in children's peer collaborations. *Developmental Psychology*, *31*(2), 207-220. doi:10.1037/0012-1649.31.2.207.

Tindale, R. S. (1989). Group vs individual information processing: The effects of outcome feedback on decision making. *Organizational Behavior and Human Decision Processes, 44*(3), 454-473. doi: 10.1016/0749-5978(89)90019-8.

Tourniaire, F., & Pulos, S. (1985). Proportional reasoning: A review of the literature. *Educational Studies in Mathematics*, 16, 181-204. doi: 10.1007/PL00020739.

Tudge, J. R. H., & Winterhoff, P. A. (1993a). Can young children benefit from collaborative problem solving? Tracing the effects of partner competence and feedback. *Social Development, 2*(3), 242 – 259. doi:10.1111/j.1467-9507.1993.tb00016.x.

Tudge, J. R. H., & Winterhoff, P. A. (1993b). Vygotsky, Piaget and Bandura: Perspectives on the relationship between the social world and cognitive development. *Human Development, 36*(2), 61-81. doi:10.1159/000277297.

Tudge, J. R. H., Winterhoff, P. A., & Hogan, M. H. (1996). The cognitive consequences of collaborative problem solving with or without feedback. *Child Development, 67*(6), 2892-2909. doi:10.1111/j.1467-8624.1996.tb01894.x.

Van Boxtel, C., van der Linden, J., & Kanselaar, G. (2000). Collaborative learning and the elaboration of conceptual knowledge. *Learning and Instruction, 10,* 311-330. doi: 10.1016/S0959-4752(00)00002-5.

Webb, N. M. (2009). The teacher's role in promoting collaborative dialogue in the classroom. *British Journal of Educational Psychology, 79*(1), 1-28. doi: 10.1348/000709908X380772.

Webb, N. M., Franke, M. L., Ing, M., Chan, A., De, T., Freund, D., & Battey, D. (2008). The role of teacher instructional practices in student collaboration. *Contemporary Educational Psychology*, *33*(3), 360-381. doi: 10.1016/j.cedpsych.2008.05.003.

Webb, N. M., Troper, J. D., & Fall, R. (1995). Constructive activity and learning in collaborative small groups. *Journal of Educational Psychology*, *87*(3), 406-423. doi:10.1037/0022-0663.87.3.406.

Zimmerman, B. J., & Martinez-Pons, M. (1990). Student differences in self-regulated learning: Relating grade, sex and giftedness to self-efficacy and strategy use. *Journal of Educational Psychology, 82*(1), 51-59. doi:10.1037/0022-0663.82.1.51.

Table 1.

*Distribution of participants to dyadic pairing conditions, based on own and partner's competence level**

Pairing condition		Feedback condition	
Dyad member 1	*Dyad member 2*	*Without Feedback*	*With Feedback*
W_1	-	$N = 22$	$N = 15$
W_1	W_1	$N = 40$	$N = 40$
W_1	W_2	$N = 44$	$N = 44$
W_1	R	$N = 36$[a]	$N = 34$[a]
W_2	-	$N = 13$	$N = 18$
W_2	W_2	$N = 42$	$N = 40$
W_2	R	$N = 30$[a]	$N = 44$[a]

* W_1 = student who employed only additive reasoning strategy on pretest, W_2 = student who used a proto-proportional strategy at least once, R = student who used a full-fledged, proportional strategy at least once.

[a] Number of participants in these cells include the R partners of W_1 and W_2 participants. Total number of participants, excluding the R partners is $N = 390$.

Table 2.

The given and the solutions of the five items of the 5BlocksTaskTest

Item	*N(A)*	*N(B)*	*N(C)*	*N(D)*	*Given weight relation between A and B*	*Weight relation to be found between C and D*
1	27	27	30	29	weight(A) > weight(B)	weight(C) > weight(D)
2	26	27	30	30	weight(A) < weight(B)	undetermined
3	26	27	30	31	weight(A) = weight(B)	weight(C) > weight(D)
4	10	9	31	28	weight(A) < weight(B)	weight(C) < weight(D)
5	10	16	24	37	weight(A) = weight(B)	weight(C) > weight(D)

Table 3.

The given and the solutions of the two Block task items in the intervention phase

Item	*N(A)*	*N(B)*	*N(C)*	*N(D)*	*Given weight relation between A and B*	*Weight relation to be found between C and D*
I1	11	10	34	31	weight (A) < weight (B)	weight (C) < weight (D)
I2	10	12	24	27	weight (A) = W(B)	weight (C) < weight (D)

Table 4.

Coding categories for students' proportional reasoning strategies

Category	*Description*	*Examples*
Visual reasoning	Judgment is based on visual features of blocks C and D or reached by counting the number of bricks in those two blocks. Blocks A and B are ignored.	''C is heavier because it has more bricks than D."
Additive reasoning strategy	The student relates to all four blocks, considering the difference between the number of bricks to establish his/her judgment. Some explanations at this level compare the difference between A and B to that between C and D; others compare the difference between A and C to that between B and D.	''The difference between A and C is that C has 3 more bricks and A is lighter. The difference between B and D is also three bricks more to D, so C will be lighter."
Proto-proportional reasoning strategy	The student relates blocks C and D to blocks A and B, considering the weight of bricks relatively to whole blocks.	''In B there are less bricks than in A, and B weighs more, so each brick in B weighs more than in A. In D there are less bricks than in C. I know that each brick in D weighs more, but there are more bricks in C than in A and in D than in B. So I don't know. I can't decide if this one brick is not too little to balance''
Proportional reasoning strategy	In addition to considering the weight of bricks relative to whole blocks, explanations at this level	''26 bricks of B weigh more than 27 bricks of A. In D there are 29 bricks and in C, 30. I know that

take into account discrepancies between blocks relative to their magnitude.	26/27 is less than 29/30, so D will be heavier.’’

Table 5.

Raw mean (and SD) pretest and posttest scores on the 5BlocksTaskTest, N = 390

		With Feedback				*No Feedback*			
Own level of competence	*Pairing condition*	*Single*	*with W_1 peer*	*with W_2 peer*	*with R peer*	*Single*	*with W_1 peer*	*with W_2 peer*	*with R peer*
W_1	Pretest	2.00 (.00)	2.00 (.00)	2.00 (.00)	2.00 (.00)	2.00 (.00)	2.00 (.00)	2.00 (.00)	2.00 (.00)
	Post-test	2.09 (.14)	2.17 (.66)	2.25 (.36)	2.74 (.72)	2.25 (.50)	2.04 (.16)	2.29 (.42)	2.33 (.36)
W_2	Pretest	2.45 (.28)	2.35 (.22)	2.37 (.21)	2.55 (.27)	2.62 (.33)	2.48 (.27)	2.50 (.28)	2.49 (.37)
	Post-test	2.78 (.52)	2.52 (.59)	2.47 (.41)	3.05 (.46)	2.71 (.59)	2.58 (.48)	2.48 (.52)	2.64 (.53)

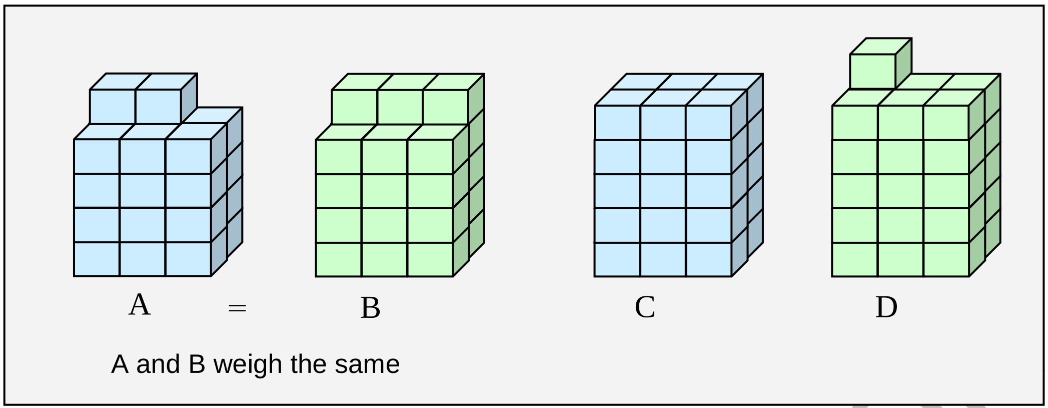
Table 6.

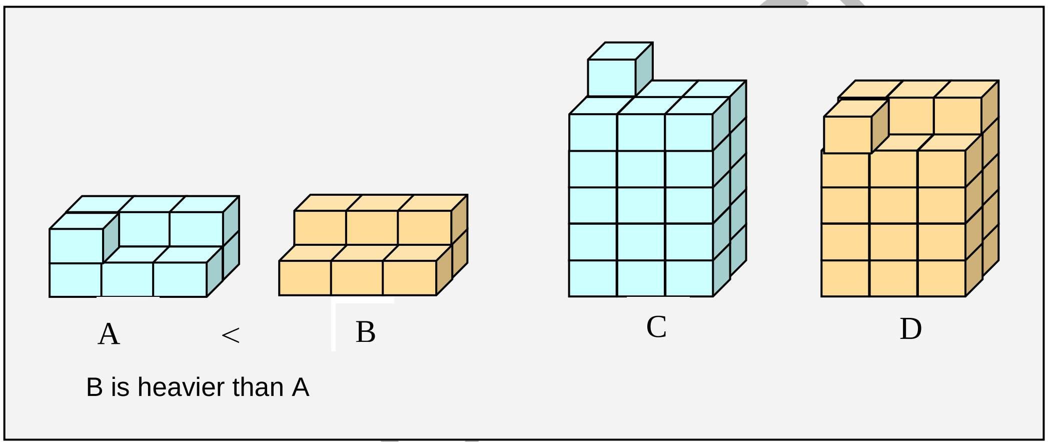
Estimates of marginal mean (and SE) learning gains of W students, by collaborative condition (dyadic or singleton) and outcome feedback condition (with or without), N = 390.

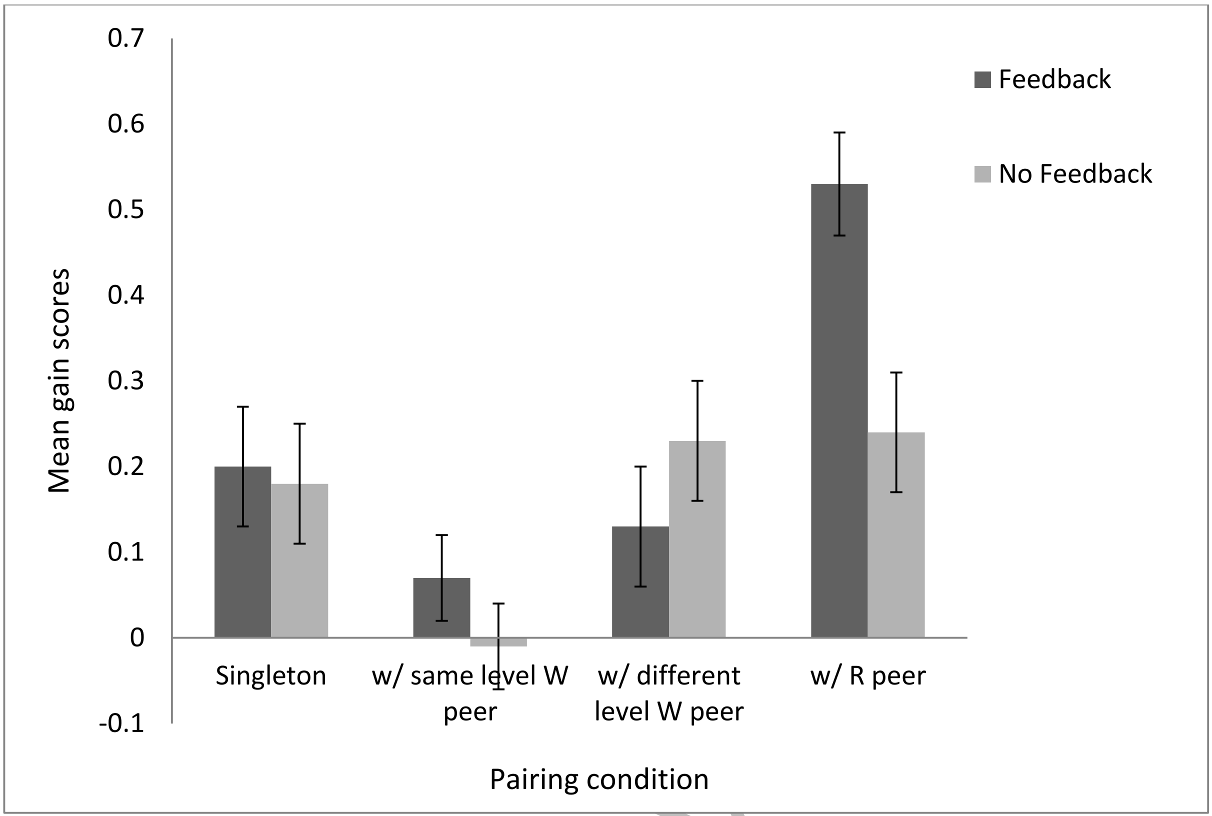
	Singleton	Dyadic	Total
Outcome feedback	.17 (.07)	.24 (.03)	.20 (.04)
No outcome feedback	.16 (.07)	.11 (.03)	.14 (.04)
Total	.17 (.04)	.18 (.02)	

Figure 1. Two items of the 5BlocksTaskTest (items 3 and 4).

Figure 2. Estimates of marginal mean (and SE) gain scores for W_x by pairing and outcome feedback condition.


A
=
B
C
D
A and B weigh the same


A
<
B
C
D
B is heavier than A


0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-0.1
Mean gain scores
Feedback
No Feedback
Singleton
w/ same level W peer
w/ different level W peer
w/ R peer
Pairing condition