UC Merced
Proceedings of the Annual Meeting of the Cognitive Science Society

Title

Conceptual change in proportional reasoning: Effects of collaboration, own / partner reasoning level and hypothesis testing

Permalink

https://escholarship.org/uc/item/2s51s2v1

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 35(35)

ISSN

1069-7977

Authors

Asterhan, Christa
Schwartz, Baruch
Cohen-Eliyahu, Noa

Publication Date

2013

Peer reviewed

Conceptual change in proportional reasoning: Effects of collaboration, own / partner reasoning level and hypothesis testing

Christa S.C Asterhan (asterhan@huji.ac.il)
Baruch B. Schwarz (msschwar@mscc.huji.ac.il)
Noa Cohen-Eliyahu (noacoe@gmail.com)
School of Education, Hebrew University of Jerusalem, Mt. Scopus
Jerusalem, 91905 Israel

Abstract

Systematic research of instruction-based conceptual change in Mathematics and Science is characterized by examining the effectiveness of a particular instructional principle in isolation. It is suggested that the field could gain from studying how different instructional principles interact when they are combined. The goal of this research was to systematically study the combined effects of collaborative learning and hypothesis testing on cognitive growth. In a randomized experiment, 496 9th graders solved challenging tasks that required fully developed proportional reasoning. Half of them were given the opportunity to test their solutions. Based on individual pretests, each student was assigned to one of three competency levels (low, medium, high), and randomly assigned to either work alone or with a (low, medium, high) peer. The findings show that the effectiveness of hypothesis testing are conditioned by fine-grained differences in the contingencies between the target student's level of competence, the peer partner's level of competence and the feedback they receive from the objective testing device.

Most of the early research on cognitive growth through peer collaboration focused on the question of optimal dyad composition (e.g., Messer, Joiner, Loveridge, Light & Littleton, 1993; Tudge & Winterhoff, 1993). However, results have overall been inconclusive and research has largely been abandoned in favor of process-oriented investigations, such as peer dialogue (e.g., Asterhan & Schwarz, 2007, 2009; Schwarz, Neuman & Biezuner, 2000) or other instructional techniques to elicit cognitive conflict, such as collaborative hypothesis testing. (e.g., Howe, Tolmie, Duchak-Tanner & Rattay, 2000; Howe, Tolmie & Rodgers, 1992).Hypothesis testing tasks require learners to translate their conceptual knowledge into hypotheses and subject these to empirical evaluation. When disconfirmed, it may confront learners with compelling evidence that they should reconsider their prior understanding even when two learners agree on their predictions (e.g., Howe et al, 2000). Vice versa, when a prediction is confirmed, it validates the explanation that led to the prediction.

In this paper, we present findings from a new study that examines whether the effects of hypothesis testing techniques depend on dyad compositions. We predict that it is. First of all, hypothesis testing in collaborating dyads may create conflict in W-W dyads (two 'wrong' learners), and settle a social conflict between members W-R dyads (one 'right' and one 'wrong' learner), who each gave different predictions and explanation. The success of hypothesis testing in socio-cognitive conflict tasks, however, hinges on a careful design: only the correct explanation or strategy should lead to a confirmation. If not, the feedback may confirm an individual's naïve, incorrect conception.

If designed carefully, this can then lead to quite powerful learning opportunities: For instance, a 'wrong' (W) student that collaborates with a 'right' (R) student will not only be exposed to a higher level of reasoning during the discussion phase, but will also receive empirical confirmation that this reasoning is correct. That is likely to be a quite powerful combination. Students in a W_x-W_x pair on the other hand, would be expected to reach quick agreement without much discussion, but shown wrong in the hypothesis testing phase, forcing them to generate a new, higher-level explanation for these findings all by themselves. Lastly, in W_x-W_y pairs the outcomes are likely to be contingent on the competency level of the particular student: A lower competency W student (W_1) is likely to benefit more from interaction with a slightly more competent W student (W_2) when there is no hypothesis testing than with it. The reason for this somewhat counterintuitive expectation is that if the W_1 student will be convinced by W_2's reasoning in the discussion phase, this solution will be proven wrong in the hypothesis testing phase. As a result, W_1 students may very well regress back to their prior level of reasoning and W_2 students may regress as well.

Very few studies have examined whether hypothesis testing techniques are more effective in collaborative or individual conditions. Two studies are particularly relevant to ours and are worth mentioning in further detail: The first is a study reported by Ellis, Klahr & Siegler (1993) that sought to investigate the effects of feedback and collaboration on 5th graders' use of mathematical rules for decimal fractions. Each of the approximately 120 pupils in this study consistently used one of two incorrect mathematical rules that were equally wrong, but qualitatively different. They were assigned to either work alone or in W_x-W_y, W_x-W_x or W_y-W_y pairs. The results demonstrated that children who had the opportunity to collaborate with a partner were more likely to use a correct rule on a posttest than children who worked alone, but only if they were given feedback during the interaction as to whether their answers were correct or not. However, dyadic composition was not found to affect children's understanding on individual tests.

Tudge, Winterhoff and Hogan (1996) also investigated the effects of feedback (hypothesis testing) and dyad composition on early elementary school children's problem solving performance on a balance beams task ($N = 83$). Children in this study either worked alone or with a partner who was equally, less, or more competent and either did or did not receive feedback on the correctness of their predictions. In direct conflict with the findings reported by Ellis et al, the presence of a partner was more effective than

working alone only when children did not receive feedback. When children received feedback, working alone was more effective than working with a partner. Similar to the Ellis et al findings, no differences were found between the different types of dyad compositions.

The findings from these two studies then lead to quite different predictions: Based on the Tudge et al findings, students may be expected to profit more from hypothesis testing when they work alone, whereas based on the Ellis et al study and findings reported by Howe et al students are expected to benefit particularly from the combination of hypothesis testing and collaboration and hypothesis testing.

The main aim of the present study is then to settle the disparate findings with regard to hypothesis testing and dyad composition in collaborative problem solving and address the following caveats in the literature. Moreover, none one of the above-mentioned studies systematically tested the effects of hypothesis testing for the *full* range of different dyad compositions that specifies the target student's *and* the partner's competence level. Finally, they did not control for nested effects of the individual within the dyad and reported findings may thus be overestimates.

The topic domain that was chosen for this study is proportional reasoning. Research suggests that students experience difficulty with proportional reasoning problems because they over-extend numerical equivalence concepts to proportional equivalence problems (e.g., Mix, Levine, & Huttenlocher, 1999; Tourniaire & Pulos, 1985). Sophisticated tests, such as the Blocks task, have been developed to serve both as instructional interventions as well as assessment tools (e.g., Schwarz & Linchevski, 2007).

Method

Participants

Eight public junior high schools from the Jerusalem and Tel Aviv metropolitan areas in Israel agreed to participate in the study. The entire 9th grade population of each school (over 600 students) completed a screening (pretest) questionnaire, to assess each student's use of problem solving strategies. Students that 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 were excused from participation in the intervention phase (see Coding section for further details). The remaining 496 9th graders (301 boys, 195 girls) used either additive (N = 196), proto-proportional (N = 194) or proportional (N = 105) reasoning strategies and participated in the intervention stage of the study. Six students did not complete the post test (2 additive and 4 pre-proportional problem solvers, respectively).

Design

Participating students within each classroom were randomly assigned to experimental condition within each group of initial level of proportional reasoning: additive (AddS), proto-proportional (ProtoS) and proportional (PropS) strategy. The basic experimental design was 2 (hypothesis testing / no hypothesis testing) * 2 (individual / dyadic work). The dyadic condition was furthermore subdivided into 5 different pairing options: AddS-AddS, AddS-ProtoS, AddS-PropS, ProtoS-PrepS and ProtoS-PropS. The entire study then included a total of 16 different experimental conditions (see Table 1 for a distribution of the participants according to conditions).

Table 1. The 16 experimental conditions of the study

Paring condition		Hypothesis testing condition	
Dyad member 1	Dyad member 2	Without HT	With HT
Adds	-	N = 22	N = 15
Adds	AddS	N = 40	N = 40
Adds	ProtoS	N = 44	N = 44
Adds	PropS	N = 36	N = 34
ProtoS	-	N = 13	N = 18
ProtoS	ProtoS	N = 42	N = 40
ProtoS	PropS	N = 30	N = 44
PropS	-	N = 15	N = 19

Tools

The task that was used for the screening, the posttest and the interaction phase is an adaptation of the Blocks task, originally developed by Harel, Behr, Lesh & Post (1992). In any given trial in the current version of the Blocks test, 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 C and A are of identical color, and so are the bricks in shapes B and D. Students are told that the weight of each brick in shapes A and C is identical, and that the same is true for each brick in B and D. At 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 two target blocks, C and D. They are given four different options to choose from (C is heavier than D, D is heavier than C, they are of the same weight, or it is impossible to determine) and are asked to base their choice with appropriate explanations (see Figure 1 for an example).

5BlocksTaskTest. Individual student's proportional reasoning level at pre- and posttests was assessed with a pen-and-paper test compiled of five Block tasks of increasing difficulty, ranging from tasks that could be solved with any strategy correctly with any strategy (e.g.,
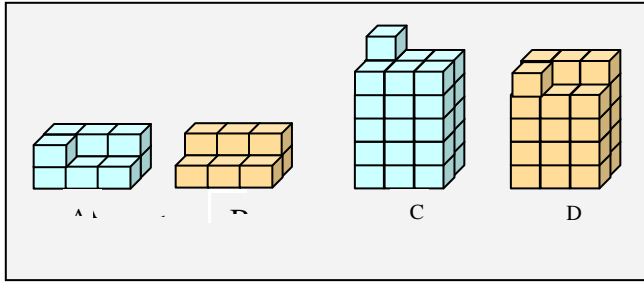
Figure 1. Example of a Blocks task item

Task 1) to tasks that could be only solved with S4 (task 4, 5).

Intervention tasks. The two items that were given during the intervention stage were not included in the 5BloksTaskTest and could only be solved correctly with proportional reasoning strategies (S4).

Coding procedures.

Students' level of proportional reasoning was assessed with the help of a slightly adapted version of a coding scheme developed by Schwarz & 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 3 different and mutually exclusive problem solving strategy categories, in ascending order of reasoning quality:

S2 (additive reasoning, grade: 2). The student takes into account the weight of a single brick in relation to the entire block, compares the target blocks to the base blocks. In this Strategy there is no multiplicative related also there is a hint to the right strategy. For example: *If A and B have the same weigh then C and D have the same weight because we add to A and B 4 bricks each to get C and D.*

S3 (proto-proportional reasoning, grade: 3). The explanation relates to all four blocks, but only refers to the nominal difference between the number of bricks of two blocks. Example: *If blocks A and B have the same weight. But there is one more bricks in B, that mean one brick in A weights more than one brick in B. so 3 bricks that added from A to C are heavier than 3 bricks that added from B to D. so C is heavier than D.*

S4 (full proportional reasoning, grade: 4). The explanation relates to all four blocks. This strategy is characterized by numerical calculation of the proportion between the four blocks. For example: *The rate between C and A is 24/10=2.4 and the rate between D and B is 37/16 =2.3125 so if they weight the same and A is multiplied in a bigger number to get C so C is heavier than D.*

Ten percent of the entire data set was coded by two independent raters, blind to condition. Inter-rater reliability was high, Cohen's κ = .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: S2 (S2 on each of the 5 pretest items), S3 (used S3 at least once, but not S4), S4 (used S4 at least once). Students that did not use at least S2 strategies on all five pretest items were excused from further participation. Performance on pretests and posttests was calculated by the mean grade of the five tasks on each test.

Procedure

All data collection and experimental interventions were completed locally in each of the 8 participating schools. Students participated in the following sequence of activities:

Stage 1: Assessment and selection. The 5BlocksTask test was administered in pen-and-paper format to all students in the participating 9th grade classes to assess their initial level of proportional reasoning and lasted between 25-40 min. Trained research assistants read aloud the instructions explaining the task. During each of the five Blocks tasks, the research assistants physically showed the 4 relevant constructions (A, B, C and D) for each task in the front of the classroom.

Stage 2: Intervention. Participating students were called to a separate room during regular school hours, in familiar rooms adjacent to 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 4 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 furthermore told that they did not have to reach consensus but that they should share ideas and explanation with each other before writing down a solution on one shared solution sheet. Students in the hypothesis condition additionally 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 may re-think [together] your solution and try to explain the outcome you received". The research assistant refrained from intervening, except to remind students of the instructions when this was 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.

Results

Analyses were conducted with a mixed model (SAS PROC MIXED) with random effects of dyad within condition and of individual within dyad and condition, on individual students' mean gains from pretest to posttest. Residuals were checked for each model separately and outliers ($z < -4$ or $z > 4$) were locally trimmed from a data set. In a few cases the kurtosis of a distribution was slightly greater than zero. When this was the case a separate analysis was conducted on the SQRT of the dependent variable (individual learning gains) and its outcomes compared to the model of its non-transformed counterpart. No differences were found in the overall pattern of results, and we therefore only report on the result from untransformed models only.

Table 3. Adjusted mean (and SE) learning gains for 'non-proportional' students
by peer pairing and hypothesis testing condition, $N = 456$.

	Pairing condition				
	Alone	Same level W partner	Different level W partner	Proportional R partner	Total
Hypothesis testing	.20 (.08)	.12 (.06)	.16 (.08)	.58 (.06)	.26 (.03)
Without hypothesis testing	.22 (.07)	-.02 (.06)	.22 (.08)	.24 (.07)	.16 (.03)
Total	.22 (.05)	.05 (.04)	.19 (.05)	.41 (.05)	

Overall effects of collaboration and hypothesis testing on learning

Table 2 presents the adjusted mean learning gains of the entire data set, according to pairing condition (working alone or in a dyad) and hypothesis testing condition (with or without weighing apparatus). A significant main effect was found for hypothesis testing, $F (1, 422) = 5.10$, $p = .024$, with students in the hypothesis testing conditions showing larger cognitive gains ($M = .25$, $SE = .04$), compare to those who did not ($M = .13$, $SE = .04$). No main effect of collaborative condition was found, $F (2, 422) < 1$, ns, and the two factors were not found to interact, $F (2, 422) = 1.48$, ns.

Table 2. Adjusted mean (and SE) learning gains for collaborative condition (dyadic or individual) and hypothesis testing condition (with or without weighing apparatus), $N = 490$.

	Individual	Dyadic	Total
With HT	.22 (.06)	.28 (.03)	.25 (.04)
Without HT	.17 (.06)	.10 (.03)	.13 (.04)
Total	.19 (.04)	.19 (.02)	

Effect of collaborating with a 'proportional' or 'non-proportional' problem solver.

We then tested whether the lack of effect for collaboration on individual learning gains could be explained by differences between students who were paired with a peer that had employed proportional strategies and students who were paired with a non-proportional peer that (i.e., either additive or proto-proportional). As in the previous model, a main effect was found for hypothesis condition, F (1, 326) = 10.40, p = .001. In addition, a main effect was found for pairing condition, F (2, 315) = 6.86, p = .001. Post-hoc analyses (with Tukey-Kramer adjustments) showed that students that collaborated with a proportional peer had larger learning gains (M = .31, SE = .04) than both students that collaborated with a non-proportional peer (M = .12, SE = .03), t (233) = 3.70, p < .001, as well as those that worked alone (M =.19, SE = .04), t (348) = 2.00, p = .046. No interaction between hypothesis testing and pairing condition was found, F (2, 315) = 1.97, ns.

Effects of dyadic pairing and hypothesis testing for non-proportional students

Next, we explored the effects of pairing and hypothesis testing amongst 'non-proportional' students only, that is: those students who had not solved any of the five pretest tasks with a full-fledged algebraic strategy. We distinguished between the following four pairing options: working without a partner (alone), being paired with a non-proportional partner of the same strategy level (same level W partner), with a partner of a different non-proportional strategy level (different level W partner) or with a partner of a full proportional strategy level (proportional R partner). Table 3 presents the adjusted mean gain scores for each of these eight conditions.

Similar to the previous models, a main effect was found for hypothesis testing, $F (1, 239) = 4.13$, $p = .043$, such that regardless of whom they were paired with, non-proportional students gained more in the weighing condition ($M = .26$, $SE = .03$) than in the non-weighing condition ($M = .16$, $SE = .03$). A main effect for pairing condition was also found, $F (2, 239) = 10.98$, $p < .001$. Post-hoc analyses (with Tukey-Kramer adjustments) showed that being paired with a proportional student ($M = .31$, $SE = .04$) resulted in larger learning gains than being paired with a same-level, 'non-proportional' peer ($p < .001$), with a different level, 'non-proportional' peer ($p = .016$) or working individually ($p = .040$).

In addition, the effect of pairing among non-proportional students was also found to be dependent on hypothesis testing condition, $F (2, 225) = 3.22$, $p = .024$. Judging from Table 3 there are two conditions that stand out in particular: The condition with hypothesis testing and a proportional partner for its comparatively high mean gain score ($M = .58$, $SE = .06$), and the condition no hypothesis testing / same-level non-proportional partner for its comparatively low mean gain score ($M = -.02$, $SE = .06$). Tukey-Kramer tests for multiple comparisons confirmed these impressions: When students were given the opportunity to test their predictions with a testing device, being paired with a 'proportional' peer indeed led to better learning gains compared to working with a same-level, 'non-proportional' peer ($p < .001$), with a different level, 'non-proportional' peer ($p = .001$) or individually ($p = .007$). There were no differences between being paired with a same-level partner, a different-level wrong partner or working alone. device, $t (240) = 3.53$, $p = .012$. Comparisons between weighing and non-weighing condition in the other three pairing conditions did not yield any significant differences. Thus, it seems that

a) Additive problem solvers
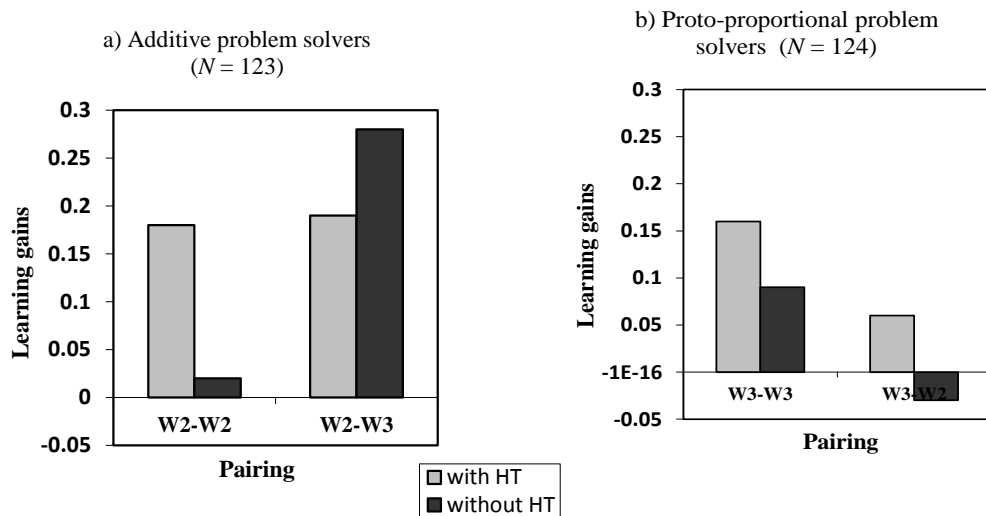(N = 123)

b) Proto-proportional problem
solvers (N = 124)

Figure 2 The effect of pairing with a non-proportional peer and hypothesis testing on learning gains, for additive (1a) and proto-proportional (1b) problem solvers

for 'non-proportional' learners as a group, neither hypothesis testing nor the pairing with a 'proportional' peer by itself resulted in learning gains, but only the combination of the two. This is further supported by the finding from post-hoc comparisons that 'non-proportional' learners in the hypothesis testing condition who were paired with a 'proportional' partner had significantly higher gains scores than students in each of the other 7 conditions.

However, when 'non-proportional' learners did not have access to a hypothesis testing device, being paired with a 'proportional' student did not have any advantage over any of the other pairing conditions (all comparisons were *ns*). Moreover, 'non-proportional' learners who are paired with a 'proportional' student gain more when they are given the opportunity to test their predictions with a hypothesis testing

Interestingly, when learners did not have access to the hypothesis testing device, 'non-proportional' students gained least when they were paired with a partner from the same level, and significantly less so than when working alone ($p = .009$), with a 'proportional' partner ($p = .005$) or with a different level 'non-proportional' partner ($p = .012$). No differences were found between the latter three conditions.

Effects of W_x-W_y pairing and hypothesis testing for different types of 'non-proportional' students

The findings reported above seem to indicate that for non-proportional students being paired with a different-level, non-proportional partner student (W_x-W_y pairing) is only preferable when students do not have access to a hypothesis-testing device, but that there is no advantage to this pairing when they have the opportunity to test their

predictions. However, these findings disregard differences in the target student's initial strategy level. The effect of W_x-W_y pairing and hypothesis testing was then separately tested for students that were initially diagnosed as 'additive' problem solvers in the Blocks task and for those that were diagnosed as 'proto-proportional' problem solvers (see Method section).

Figure 2 presents the adjusted mean learning gains for additive (Fig 2a) and for proto-proportional problem solvers (Fig 2b) that are paired with non-proportional peers. In contrast to the previous models, no main effects for hypothesis testing were found, neither for additive problem solvers ($F < 1$), nor for proto-proportional learners ($F < 1$). Among additive problem solvers, a main effect was found for pairing condition, $F (1, 56.7) = 6.01$, $p < .017$, with students who were paired with a proto-proportional peer showing higher learning gains ($M = .24$, $SE = .03$) compared to those that were paired with a same-level peer ($M = .10$, $SE = 03$). Pairing condition was also found to interact with hypothesis testing, $F (1, 56.7) = 5.56$, $p < .022$. Post-hoc analyses (with Tukey-Kramer adjustments) showed that when paired with another additive problem solver, they learning gains were higher with hypothesis testing ($M = .18$, $SE = 05$), than without it ($M = .02$, $SE = 05$), $t (55) = 2.39$, $p = .032$. When they were paired with a proto-proportional peer, on the other hand, additive problem solvers seemed to gain more without the hypothesis device ($M = .28$, $SE = 06$) than with it ($M = .19$, $SE = 06$). This apparent difference did not reach statistical significance however, $t (73) = 1.12$, *ns*.

For the proto-proportional problem solvers, on the other hand, no effect were found for neither pairing condition ($F (81) = 1.34$, *ns*), hypothesis testing ($F < 1$), nor their interaction ($F < 1$).

Discussion

Previous studies have examined the effects of hypothesis testing and collaborative learning on cognitive growth (e.g., Ellis et al, 1993; Howe et al, 2000; Schwarz et al, 2000; Tudge et al, 1996). Unfortunately, this literature has yielded a mixed pattern of results. In the present study we revisited the major research questions in this field with a controlled experimental design that systematically explored the full range of dyadic compositions and with statistical models that controlled for nested effects. Overall, the findings show that the answer to the question whether hypothesis testing-based interventions for learning are more effective in individual or collaborative settings, really depends on the level of analysis and the comparisons being made.

First of all, when all different types of dyadic compositions are included in the data set but not further specified, hypothesis testing was overall found to improve students' learning gains. This finding is consistent with earlier research on the effectiveness of providing students with feedback that consistently confirms correct predictions and disconfirms predictions based on incorrect understanding (e.g., Tudge & Winterhoff, 1993; Tudge et al, 1996). Collaboration, on the other hand, was not found to have an overall advantage over individual work. It was neither found to improve learning through hypothesis testing as is often expected and as found in other studies (Ellis et al, 1993). It is often believed that peer collaboration allows learners to discuss different explanations and generate interpretations of the hypothesis testing outcomes (Howe et al 2000). However, such potential benefits of collaboration are not detectable when the full range of dyadic pairings are included but not further specified.

A further dissection of the general construct of 'collaboration' according to the target student's *and* the partner's competence levels uncovers that interaction with a more competent peer only improves learning under certain specific conditions: For non-proportional ("wrong") students, the combination of hypothesis testing and being paired with a proportional ("right") partner was particularly powerful. However, similar to Ellis et al (1993) we found that when students received no feedback from the equipment (no hypothesis testing), singletons, students paired with proportional peers and students paired with different level non-proportional peers showed only comparable (moderate-to-small) gains. In concordance with previous findings (Ames & Murray, 1983; Schwarz et al, 2000), students who were paired with a same-level "wrong" peer without the opportunity to receive any feedback through hypothesis testing did not improve at all. The pattern that emerges from these findings seems to underline the importance of the combination of exposure to higher-order reasoning strategies *and* the confirmation of the correctness of these strategies by an objective test. This is not an additive effect, since neither the exposure to higher-order reasoning strategies, nor the conflict created by the disconfirmation of incorrect predictions alone led to substantive learning gains.

This subtle contingency of, on the hand, the kind of feedback that is obtained from objective testing and, on the other, the persuasiveness of a higher-order reasoning strategy becomes even more evident when we considered the wrong-wrong pairings only: The benefits of interaction with a more competent peer and hypothesis testing were found to hold only when the test proved that the predictions of this more competent peer were correct. When less competent (using additive strategies) interacted with a more competent peers (using proto-proportional strategies), the former actually gained more *without* hypothesis testing. When there is no opportunity to test the correctness of predictions, the verbal explanation provided by the slightly more competent peer may convince the lower competence peer to the more sophisticated reasoning strategy, thus improving their performance on posttests. However, with access to hypothesis testing devices, the predictions of the more competent peer will be disconfirmed, and with it the (slightly) more sophisticated reasoning strategy.

References

Asterhan, C. S. C. & Schwarz, B. B. (2009). The role of argumentation and explanation in conceptual change: Indications from protocol analyses of peer-to-peer dialogue. *Cognitive Science, 33*, 373-399.

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

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

Howe, C., Tolmie, A., & Rodgers, C. (1992). The acquisition of conceptual knowledge in science by primary school children: Group interaction and the understanding of motion down an incline. *British Journal of Developmental Psychology, 10*, 113–130.

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 & Instruction, 10*, 361-391.

Messer, D. J. Joiner, R. Loveridge, N. Light, P. & Littleton, K (1993). Influences on teh effectiveness of peer interaction: children's level of cognitive development and teh relative abiluty of partners. Social Development, 2, 279 – 294.

Mix, K. S., Levine, S. C., & Huttenlocher, J. (1999). Early fraction calculation ability. *Developmental Psychology, 35*, 164-174.

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. *, 17*(5), 510-531.

Tudge, J. R. H., Winterhoff, P. A. and Hogan, D. M. (1996). The Cognitive Consequences of Collaborative Problem Solving with and without Feedback. *Child Development, 67*, 2892-2909