Superficial, rather than true, knowledge interdependence in collaborative learning fosters individual knowledge transfer

Anne Deiglmayr (anne.deiglmayr@ifv.gess.ethz.ch)
ETH Zurich, Research on Learning and Instruction, Universitaetsstrasse 41
8092 Zurich, Switzerland

Lennart Schalk (lennart.schalk@ifv.gess.ethz.ch)
ETH Zurich, Research on Learning and Instruction, Universitaetsstrasse 41
8092 Zurich, Switzerland

Abstract

We test the hypothesis that *superficial knowledge interdependence* is more effective in fostering individual learning from collaboration than the *true knowledge interdependence* often realized by jigsaw-type collaboration arrangements. Based on research on group information-processing, we argue for the benefits of distributing only contextual information, but not core principles between learners, establishing superficial knowledge interdependence. In a computer-supported collaborative learning environment, 78 university students learned about stochastic urn models. Knowledge interdependence was established by systematically distributing learning materials within student triads, so that students either became experts for an urn model, establishing true knowledge interdependence, or for one of the embedding cover stories, establishing superficial knowledge interdependence. Afterwards, all triads worked on the same collaborative tasks, and were exposed to all models. Results show successful learning across conditions, but superior knowledge transfer in triads collaborating under superficial knowledge interdependence. Benefits were highest for low prior knowledge learners.

Keywords: computer-supported collaborative learning; learning through comparison; knowledge interdependence; knowledge transfer

Introduction

In this paper, we explore different ways of distributing information between collaborative learners, with the goal of promoting the interactive construction of mathematical principles during learning from collaborative comparison of worked examples. In doing so, we address the more fundamental question of what characterizes optimal *knowledge interdependence* in collaborative learning, as assessed by measures of individual learning and transfer.

Collaborative learning has the potential of engaging students in forms of interactive knowledge construction that yield learning outcomes beyond those within the reach of an individual learner (Chi, 2009). However, this requires a certain amount of *knowledge interdependence* between students, that is, the individual students should hold a certain amount of *unshared (unique) knowledge*, ideas, and perspectives. The deliberate creation of knowledge interdependence is an important factor in many instructional methods for fostering collaborative learning, with the jigsaw collaboration script as their prototype. In a jigsaw collaboration script, each learner becomes an expert for a specific domain before collaborating with other learners who have studied a different domain. To ensure fruitful collaboration, the distribution of expertise within groups typically ensures that "none of the group members has enough information or knowledge to solve the task alone" (Dillenbourg & Jermann, 2007, p. 292), establishing *true knowledge interdependence*.

In fact, differences in prior knowledge and perspectives can lead to fruitful knowledge co-construction, in which ideas are critically evaluated, knowledge is elaborated and restructured, and more abstract representations are derived (Andriessen, Baker, & Suthers, 2003; Schwartz, 1995). When learners integrate and transform their complementary knowledge resources, new knowledge can be created that no individual learner would have been capable of constructing (Deiglmayr & Spada, 2011). On the other hand, research on group information processing shows that much of students` unshared knowledge remains unshared in real group discussions. For example, Buchs, Butera, and Mugny (2004) showed that students studying with a jigsaw collaboration script learned substantially less about their partner's domain of expertise than about their own, even though they were instructed to teach one another during a face-to-face learning phase. Deiglmayr and Spada (2011) showed that students had severe difficulties integrating interdependent information that was distributed between them.

Educators face the challenge of creating knowledge interdependence in a way that ensures that learners' discussions, and the cognitive activities involved, are focused on the most relevant learning content. Establishing true knowledge interdependence, as in classical jigsaw-type collaboration scripts, may not always be the optimal way to achieve this goal. Rather, we argue that *superficial knowledge interdependence* is often the better solution. Superficial knowledge interdependence denotes that *core structures*, such as domain principles and important concepts, remain shared between learners, while only *contextual information*, such as illustrative examples or application contexts, is distributed between learners. The fact that all relevant structural information is given to all students from the beginning maximizes the chance that each learner becomes familiar with the relevant principles via constructive learning processes, while the distributed context information still creates sufficient interdependence for fostering truly interactive knowledge construction (Chi, 2009). In this paper, we test this *"shared structure, distributed context"-hypothesis* in a schema-abstraction learning setting (learning by collaborative comparison), with a learning domain that allows for a straightforward distinction between structure and context (word problems instantiating mathematical principles within different application contexts).

Learning by collaborative comparison

Comparing and contrasting worked examples has proven an efficient way of fostering learning and transfer (for recent reviews see Alfieri, Nokes, & Schunn, in press, and Rittle-Johnson & Star, 2012). According to this approach at least two carefully constructed worked examples, which are instantiation of the to-be-learned principle or schema, are presented simultaneously in space and time. Learners are prompted to compare and contrast the examples in order to identify commonalities and differences (e.g., Gentner, Loewenstein & Thompson, 2003; Schalk, Saalbach, & Stern, 2011). These activities require learners to map and structurally align aspects of the worked examples, which "leads to learning via abstraction, rerepresentation, inference-projection, and difference-detection" (Gentner, 2010, p. 753). These are higher-order learning processes in which learners need to focus on deep, structural information rather than on contextual features, and to elaborate the to-be-learnt principles. In our collaborative comparison script, students begin with slightly different sets of examples from which they have to generate joint explanations of principles. This presumably fosters principle-based comparisons and elaboration via processes of grounding (Andriessen et al., 2003; Schwartz, 1995) and knowledge co-construction (Chi, 2009). Because the to-be-learnt principles (structural information) are embedded within different cover stories (contextual information), collaborative comparison as an instructional method allows to design well-controlled tests of the *"shared structure, distributed context"-hypothesis*.

The domain: Learning to reason with probability

The relevant principles that students could learn in our experiment were *urn models*. These models serve to describe the probability of a series of random events (i.e., multilevel random experiments) in basic probability theory and allow for differentiating precisely between structure (urn models and the principles underlying them) and context (application contexts in the form of story problems).

A sound understanding of basic probability theory is a fundamental precondition for acquiring the ability to solve problems in statistics and, as such, is required in many professions and academic disciplines. High quality teaching seems to be particularly important as reasoning about probabilities does not come naturally to most people, and biases and misconceptions are abundant (Kahneman, Slovic, & Tversky, 1982). Basic principles of probability theory and stochastics are introduced quite early in high school mathematics. In Switzerland, for example, the principles governing multilevel random events (the learning domain from which our learning materials were taken) is introduced as early as in eighth grade. Typical problems are, for example, finding the probability of getting twelve points when throwing two dice, or finding the likelihood of guessing the right combination of numbers in a lotto game. The ultimate goal is that mathematical/statistical knowledge acquired in school will be applied outside the classroom and in students' later work; that is, that transfer occurs (Singley & Anderson, 1989). However, transfer does not come about naturally even for these basic probability theory principles, and even university students have difficulties with basic stochastic concepts (Gal, 2002).

The present research

In our experiment, university students had the chance to refresh and deepen their knowledge about basic probability theory, specifically, their knowledge about multilevel random events. The most important conceptual knowledge learners need to acquire when learning about multilevel random events is the ability to differentiate between four different *urn models*, in which random events are modeled as balls being drawn from an urn.

We combined learning through collaboration with learning triggered by comparing and contrasting worked examples in a *collaborative comparison script.* The script was modeled after a prototypal jigsaw script with an individual and a collaborative learning phase, implemented within a computer supported collaborative learning (CSCL) environment. Learning materials consisted of worked examples, which embedded the to-be-learned urn models in different cover stories. We varied whether, prior to collaboration, students became experts for one urn model (MODEL experts: *true knowledge interdependence*) or for one cover story (STORY experts: *superficial knowledge interdependence*). This setting allowed us to test our hypothesis that superficial knowledge interdependence would be more effective than true knowledge interdependence in fostering students` learning.

Method

Participants

Participants were 87 students of universities in Zurich (Switzerland), majoring in a wide range of subjects (students of mathematics or statistics were excluded). All participants spoke German or Swiss German as a native language. They were paid for participation. Participants were randomly assigned to triads and conditions. We excluded three triads from analysis because at least one of their members did not pass the threshold of four out of six correct answers in a basic prior knowledge test. This test assessed basic skills necessary for learning about multilevel random events (e.g. finding the likelihood of single random events in story problems; adding and multiplying fractions), or because they did not follow instructions. These exclusion criteria left a total of 78 participants (42 female, 33 male) in 26 triads. Their age ranged from 18 to 36 years (M = 24.4, SD = 4.0).

Materials

Four urn models from probability theory (specifically, multilevel random events) were the core learning content of our learning environment. These four models result from combinations of two principles: *relevance of order* (the order in which balls are drawn from an urn is relevant vs. irrelevant) and *replacement* (the balls are drawn with replacement vs. without replacement). We will refer to these four models as *Model 1* (order relevant, without replacement), *Model 2* (order relevant, with replacement), *Model 3* (order irrelevant, without replacement), and *Model 4* (order irrelevant, with replacement). Story problems exemplified the four urn models by embedding them in simple cover stories (see Table 1 for examples). We used three different story problems, adapted with modifications from Berthold and Renkl (2009). In the remainder of this paper, these stories will be referred to as *Story 1* (random events = the distribution of bicycle helmets among participants in a biking course), *Story 2* (random events = ranking results in a competition among equally capable ski-jumpers), and *Story 3* (random events = the drawing of unlabeled gas bottles from cupboards in a chemist`s lab). In the learning materials, we used nine story problems that result from crossing *Models 1-3* with *Stories 1-3*. They were presented in the form of worked examples, that is, together with an arithmetic solution approach and a final numerical solution (as in Table 1). The three problems resulting from crossing *Model 4* with *Stories 1-3* were used as transfer tasks in the post test. All materials were presented within a computer-based learning environment.

Measures and Scoring

Pretest In addition to the six basic knowledge questions used for screening participants (see Participants), the pretest contained four story problems assessing learners' prior knowledge about Models 1-4. The cover stories differed from those used in the learning phase. For each problem, one point could be obtained for generating an equation that corresponded to the model underlying the story problem.

Posttest The posttest had three sections. Within each section, the order of tasks was randomized. For each task, one point could be obtained for generating an equation that corresponded to the correct model. In the first section, three *familiar tasks* represented *Models 1-3,* each embedded in one of the *Stories 1-3* that students already knew from the learning environment, but with new numerical values. In the second section, six *direct application tasks* embedded *Models 1-3* in novel cover stories (two tasks for each model). The third section comprised the three tasks that result from crossing *Model 4* with *Stories 1-3*. These *Model 4 transfer tasks* were included to measure transfer of the principles underlying *Models 1-3*: Since the four urn models result from crossing the principles *relevance of order* (relevant / irrelevant) and *replacement* (with / without), the fourth model can be derived from the other three. Students were told that the transfer tasks constituted a new type of model, but that they would be able to solve them by combining what they had learned during the learning phase.

Procedure

Students came to our lab in groups of up to 18 participants. After a brief introduction, they were randomly assigned to computer work stations. Each student sat in his or her own cubicle, so that there was no face-to-face contact possible between learners. Students did not know with whom they

Table 1: Three worked examples from the learning materials (translated from the original language, German) exemplifying the three models and the three cover stories used in the learning phases

Model 1, Story 1	*Model 2, Story 2*	*Model 3, Story 3*
You and your friend participate in a two day mountain bike course. Each day, the instructor brings five bicycle helmets in five different colors which are randomly distributed among the course participants in the morning, and collected again in the evening. On both days, you are the first to receive a helmet, and your friend is the second. What is the probability for you to get the red helmet on the first day and the yellow helmet on the second day?	The four ski jumpers Adam, Beat, Christoph, and Daniel test a newly build ski-jumping hill today. The four ski jumpers have all performed equally well on previous competitions, thus, it only depends on random factors (e.g., wind regime) which of them will jump the greatest distance. There are two rounds of jumps. What is the probability that Adam will be on the first rank and Daniel on the second rank after the first round of jumps?	A chemist stores noble gases in two safes. There are the same three noble gases (argon, krypton, and xenon), in three identical single bottles, in both safes. Unfortunately, her colleague forgot to label the bottles. For her experiments, the chemist needs two different gases. The chemist takes one bottle out of each safe. What is the probability for her to obtain one bottle of argon and one of xenon?
Approach Solution $\frac{1}{5} * \frac{1}{5} \quad = \frac{1}{25}$	Approach Solution $\frac{1}{4} * \frac{1}{3} \quad = \frac{1}{12}$	Approach Solution $\frac{1}{3} * \frac{1}{3} + \frac{1}{3} * \frac{1}{3} \quad = \frac{2}{9}$

were collaborating, and were logged into the system with an anonymous, gender-neutral nickname. After arriving at their workstations, students filled in a questionnaire on demographic variables and worked on the pretest individually. Afterwards, and before starting the learning phase, students received an introducing to the chat tool, and the three students who had been assigned to the same triad engaged in a brief warming-up chat session. The learning phase was segmented into an individual learning phase followed by a collaborative learning phase. Table 2 gives an overview of the worked examples presented in both phases, along with the self-explanation prompts provided (abbreviated for the individual learning phase).

Table 2: Learning materials (worked examples) for both experimental conditions. Worked examples are denoted by their combination of *Model* (M1-3) and *Story* (S1-3).

	Individual learning phase	
	MODEL-experts	STORY-experts
Learner A	M1S1-M1S2-M1S3 *Commonalities?* *Differences?*	M1S1-M2S1-M3S1 *Commonalities?* *Differences?*
Learner B	M2S1-M2S2-M2S3 *Commonalities?* *Differences?*	M1S2-M2S2-M3S2 *Commonalities?* *Differences?*
Learner C	M3S1-M3S2-M3S3 *Commonalities?* *Differences?*	M1S3-M2S3-M3S3 *Commonalities?* *Differences?*

	Collaborative learning phase
Triad A B C	Screen 1: M1S1-M1S2-M1S3 *Why are the fractions multiplied rather than added up?*
Triad A B C	Screen 2: M2S1-M2S2-M2S3 *Why is the fractions' denominator decreasing?*
Triad A B C	Screen 3: M3S1-M3S2-M3S3 *Why does the solution require both addition and multiplication?*

The experimental variation was established in the individual learning phase, in which each learner studied three worked examples that were presented side-by-side on one screen. Learners were prompted to compare the examples and to list the most important similarities and the most important differences. Each member of a triad was assigned a different set of examples, so that, among them, the three learners studied all nine examples that result from crossing *Models 1-3* with *Stories 1-3*. In the MODEL-experts condition, each triad member became an expert for a different urn model (true knowledge interdependence), whereas in the STORY-experts condition, each triad member became an expert for a different cover story (superficial knowledge interdependence).

In the collaborative learning phase, materials and instructions were identical for all triads, regardless of experimental condition. Three sets of worked examples, corresponding to *Models 1-3*, were presented on three consecutive screens (see Table 2). Thus, each and every learner was exposed to all nine worked examples during the collaborative learning phase. The triads compared and contrasted the worked examples and generated collaborative self-explanations. For each set of worked examples they were prompted to focus on one specific feature of the urn model being exemplified (see Table 2 for details). Triads used the chat tool in order to discuss their answer. Once group members had agreed on a joint solution, they went on to the next screen. After the collaborative learning phase, students worked on the posttest individually. All in all, the experiment took about 100 minutes.

Results

There were no relevant differences between experimental conditions in participants` age, final high school math grade, or performance on the basic knowledge test used for participant screening (all *ts* < |1.5|; all *ps* > .15). Further, conditions did not differ significantly in the proportion of females/males ($\chi^2_{(df = 1)}$ = .83; p = .36). Conditions also did not differ in the distribution of students who solved 0, 1, 2, 3 or 4 of the *pretest Models 1-4 tasks* correctly ($\chi^2_{(df = 3)}$ = .42; p = .94) indicating similar levels of prior knowledge (see Table 3 for mean proportions correct).

Table 3: Mean proportions correct (and standard deviations) of pre- and post-test scores (total N = 78)

	MODEL Experts	STORY Experts	whole sample
pretest:			
Models 1-4 total	.55 (.24)	.54 (.26)	.55 (.25)
Models 1-3 only	.68 (.25)	.66 (.25)	.67 (.25)
Model 4 only	.18 (.39)	.21 (.41)	.19 (.39)
posttest:			
Models 1-3 familiar	.76 (.26)	.79 (.24)	.78 (.25)
*Models 1-3*application	.75 (.23)	.76 (.19)	.75 (.21)
Models 1-3 combined	.75 (.22)	.77 (.19)	.76 (.20)
Model 4 transfer	.46 (.44)	.62 (.35)	.54 (.40)

Before analyzing the post-test scores, we calculated intra-class correlations for the members of each triad in order to test for a possible hierarchical data structure. In no case was the *ICC* above .05 (all *Fs* < 1.1; all *ps* > .40), indicating only unsystematic agreement in post-test scores between triad members and, thus, a non-hierarchical data structure. Therefore, we calculated all further analyses on the level of individual learners (N = 78). Given that our data is made up by series of 0 vs. 1 (correct vs. incorrect) responses, we calculated generalized logit regression models (using SPSS`s GENLIN procedure, with a logit link function) rather than *t*-tests or ANOVAs (Jaeger, 2008). However, for ease of comparison, Table 3 gives the scores that students in the two experimental conditions obtained as mean proportions correct. Students in both conditions achieved very similar scores on the *Models 1-3 familiar* and the *Models 1-3 direct application* tasks. The differences between these two post-test sections (as within-subjects factor), experimental condition (as between-subjects factor), and their interaction were all statistically non-significant in a generalized logistic regression (all *Wald*-$\chi^2_{(df = 1)} < 2.6$; all *p*s > .11). We therefore formed a combined posttest score (Table 3: *Models 1-3 combined*).

We first looked at students` posttest performance on tasks representing Models 1-3, that is, the learning content we directly taught. Table 3 shows that students in both conditions showed an overall gain in their performance from pre- to posttest. We calculated a generalized logistic regression with solution rate as the dependent variable, time (pretest: *Models 1-3* vs. posttest: *Models 1-3 combined*) as within-subjects factor, and experimental condition as between-subjects factors. Only the effect of time was significant (*Wald*-$\chi^2_{(df = 1)} = 6.5$; $p = .01$). These findings indicate that both conditions were effective in improving the recognition and application of the three urn models that were directly taught.

On the *Model 4 transfer* tasks, however, students` posttest performance was notably higher in the STORY-experts condition (Table 3). Figure 1 shows that the absolute solution rate shows a U-shaped distribution in the MODEL-experts conditions, while the mode of the distribution in the STORY-experts condition is at the highest end of the distribution. This difference in distribution of scores between conditions is statistically significant ($\chi^2_{(df = 3)} = 8.55$; $p = .04$).

To further scrutinize the differential effects on transfer in both conditions, we took students' prior knowledge into account. We tested the effects of experimental condition, prior knowledge (specified as a covariate), and their interaction, on the number of correctly solved *Model 4 transfer* tasks in a generalized logistic regression model. We chose the combined pretest score for *Models 1-4* as the most reliable and most informative predictor; however, analyses with performance on only the items for *Models 1-3* yielded the same pattern of results; the same was true when the 15 students who had already mastered the *Model 4* task in the pretest were excluded from analysis. All postulated predictors in the model (experimental condition, prior knowledge, and their interaction) were shown to significantly predict performance on the *Model 4 transfer* tasks (for parameter estimates see Table 4; overall model likelihood ratio: $\chi^2_{(df = 3)} = 39.83$; $p < .001$). The significant interaction indicates that learners low in prior knowledge profited more in the STORY-experts condition than in the MODEL-experts condition: Prior knowledge showed a significant, positive correlation with transfer performance in the MODEL-experts condition (Spearman`s $r = .56$, $p < .001$) but a smaller, statistically non-significant correlation in the STORY-experts condition (Spearman`s $r = .25$, $p = .13$).

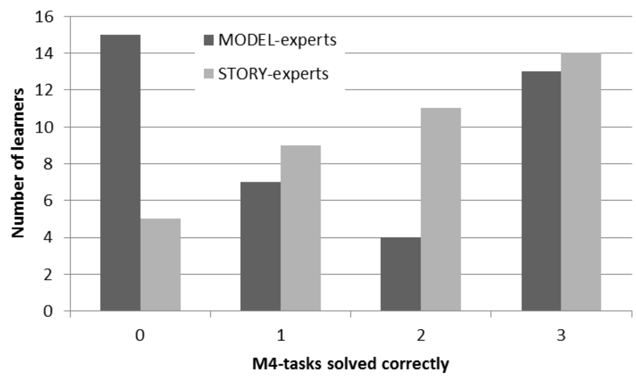

Figure 1: Distribution of learners (by experimental condition) who solved 0, 1, 2, or all 3 *Model 4 transfer* tasks correctly

Discussion

In the present study, we aimed at testing the hypothesis that in collaborative learning settings *superficial knowledge interdependence* is more effective in fostering individual learning than true knowledge interdependence. Specifically, we tested whether collaborative learning supported by a jigsaw-type collaboration script is more effective when the knowledge interdependence established between students ensures that the to-be-learnt, structural information (in our case, the three urn models) is shared from the beginning, while only contextual information (in our case, the cover

Table 4: Summary of effects in the generalized logit model with experimental condition, prior knowledge, and their interaction as predictors of students' performance on the *Model 4 transfer* tasks ($N_{subjects\_x\_trials} = 234$)

Predictor	Coefficient (*B*)	*SE*	Coefficient (*B*): *95%-CI* (Wald)	e^B	Wald $\chi^2_{(df = 1)}$	*p*
Intercept	-.37	.46	[-1.28; .54]	.69	.64	.42
Experimental Condition MODEL-experts = 0 STORY-experts = 1	-2.44	.76	[-3.94; -.95]	.09	10.29	< .01
Prior Knowledge (*Models 1-4* pretest score)	.41	.20	[.01; .81]	1.51	4.11	.04
Interaction: *Experimental Condition* x *Prior Knowledge*	.79	.33	[.15; 1.44]	2.21	5.76	.02

stories) is distributed between learners (*shared structure - distributed context hypothesis*).

The results partially support our hypothesis: Students in the STORY-experts condition (superficial knowledge interdependence) did profit more from our CSCL learning environment than students in the MODEL-experts condition (true knowledge interdependence), but only on the transfer tasks. In both conditions, students gained to a similar degree from pre- to post-test for the three models that had been trained. Since students in both conditions learned with a highly structured learning environment and with carefully constructed worked examples, this finding is reassuring. Still, STORY-experts outperformed MODEL-experts on the transfer tasks, which required them to combine the principles behind the three trained models in order to derive a solution for a fourth model that had not been introduced within the learning environment. Learners with low prior knowledge profited particularly from the superficial knowledge interdependence realized in the STORY-experts condition, that is, they were more likely to obtain a high score on the transfer tasks in this condition.

We assume that these effects arise because the superficial knowledge interdependence realized in the STORY-experts condition (1) ensures that each learner becomes familiar with all relevant principles via constructive learning processes already during the preparatory individual learning phase, while (2) the distributed context information still creates sufficient interdependence for fostering truly interactive knowledge construction (Chi, 2009). However, further fine grained analyses of individual learning (e.g. self-explanations during individual learning phase) and of collaborative processes (e.g., discourse analyses of chats) are needed to be able to precisely identify the underlying cognitive and interactive processes. Analyses currently under way include coding the quality of students` self-explanations, as a measure of the level of expertise they gained during the individual learning phase, as well as analyses of the patterns of contributions, both qualitatively and quantitatively, made within story expert and model expert triads. Further experiments will include additional test and transfer tasks in order to increase the reliability of the pre- and post-test measures, and will be designed to enable direct comparisons with purely individual (constructive) learning conditions.

Acknowledgments

(1) This research was partially supported by an early career scholarship to the first author from the Swiss National Science Foundation. (2) Both authors contributed equally.

References

Alfieri, L., Nokes-Malach, T. J., & Schunn, C. D. (in press). Learning through case comparisons: A meta-analytic review. *Educational Psychologist*.

Andriessen, J., Baker, M., & Suthers, D. (Eds.) (2003). *Arguing to learn: Confronting cognitions in computer-supported collaborative learning environments.* Dordrecht: Kluwer.

Berthold, K., & Renkl, A. (2009). Instructional aids to support a conceptual understanding of multiple representations. *Journal of Educational Psychology, 101(1)*, 70-87.

Buchs, C., Butera, F., & Mugny, G. (2004). Resource interdependence, student interactions and performance in cooperative learning. *Journal of Experimental Educational Psychology, 24(3)*, 291-314.

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

Deiglmayr, A., & Spada, H. (2011). Training for fostering knowledge co-construction from collaborative inference-drawing. *Learning and Instruction, 21*, 441-451.

Dillenbourg, P., & Jermann, P. (2007). Designing integrative scripts. In F. Fischer, H. Mandl, J. Haake, & I. Kollar (Eds.), *Scripting computer-supported collaborative learning -- Cognitive, computational, and educational perspectives* (pp. 275–301). New York: Springer Computer-Supported Collaborative Learning Series.

Gal, I. (2002). Adults' statistical literacy: Meanings, components, responsibilities. *International Statistical Review, 70(1)*, 1-25.

Gentner, D. (2010). Bootstrapping the mind: Analogical processes and symbol systems. *Cognitive Science, 34*, 752-775.

Gentner, D., Loewenstein, J., & Thompson, L. (2003). Learning and transfer: A general role for analogical encoding. *Journal of Educational Psychology, 95(2)*, 393-408.

Jaeger, T. F. (2008). Categorical data analysis: away from ANOVAs (transformation or not) and towards Logit Mixed Models. *Journal of Memory and Language, 59*, 434-446.

Kahneman, D., Slovic, V. M., & Tversky, A. (Eds.) (1982). *Judgment under Uncertainty: Heuristics and Biases*. New York: Cambridge University Press.

Rittle-Johnson, B., & Star, J. R. (2012). The power of comparison in learning and instruction: Learning outcomes supported by different types of comparison. In B. Ross & J. Mestre (Eds.), *Psychology of Learning and Motivation: Cognition in Education* (Vol. 55). San Diego: Elsevier.

Schalk, L., Saalbach, H., & Stern, E. (2011). Designing learning materials to foster transfer of principles. In L. Carlson, C. Hölscher, & T. Shipley (Eds.), *Proceedings of the 33rd Annual Conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.

Schwartz, D. L. (1995). The emergence of abstract representations in dyad problem solving. *Cognition and Instruction, 22(2)*, 129-184.

Singley, M. K., & Anderson, J. R. (1989). *Transfer of Cognitive Skill.* Cambridge, MA: Harvard University Press.