Title

Expert Blind Spot in Pre-Service and In-Service Mathematics Teachers: Task Design moderates Overestimation of Novices' Performance

Permalink

https://escholarship.org/uc/item/8qw4x3hp

Journal

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

ISSN

1069-7977

Authors

Hellman, Katharina
Nuckles, Matthias

Publication Date

2013

Peer reviewed

Expert Blind Spot in Pre-Service and In-Service Mathematics Teachers: Task Design moderates Overestimation of Novices' Performance

Katharina Hellmann (katharina.hellmann@ezw.uni-freiburg.de)
Matthias Nückles (matthias.nueckles@ezw.uni-freiburg.de)
University of Freiburg, Department of Educational Science, Instructional and School Research
Rempartstraße 11, 79098 Freiburg, Germany

Abstract

To act efficiently in the classroom, teachers need to be able to judge the difficulty of problems from a novice's perspective. However, research suggests that experts use their own knowledge as an anchor, adjust estimations for others to their own knowledge and thus underestimate the difficulty that a problem may impose on novices. Similarly, experts should underestimate the benefit for novices of task designs derived from Cognitive Load Theory (CLT), as – following the expertise reversal effect – these should be rather disadvantageous for experts. We investigated pre-service and in-service teachers' competencies in estimating the difficulty of mathematical tasks for novices. Thirty-four pre-service teachers and thirteen experienced teachers solved tasks that varied in instructional design (optimized for novices following CLT versus non-optimized). Participants solved each task and then estimated how many students of a fictional 9th grade class would be able to solve that task. Solution frequencies were collected from fifty-two 9th grade students. In both expert groups, overestimation was clearly more pronounced for non-optimized than optimized tasks, suggesting an expert blind spot that can be explained in terms of an expertise-reversal effect. The experts failed to adequately take into account the benefits of didactic task variation for novice learners. However, whereas pre-service teachers' overestimations of student performance were large and significant both for non-optimized and optimized tasks, in-service teachers' overestimations were generally small and failed to approach statistical significance. In contrast to pre-service teachers, in-service teachers seem to have a better mental model of what a student is able to achieve, thus making better judgments of student performance.

Keywords: expert blind spot; perspective taking; expertise reversal effect.

Theoretical Background

Expert Blind Spot

Peoples' judgements of others are very often based on their self-assessment and are therefore cognitively biased (e.g. Tversky & Kahneman, 1974). In line with this, research in the area of expertise has repeatedly shown that experts tend to misjudge novices' knowledge, achievement, or time on task, amongst others, to a certain degree (e.g. Herppich, Wittwer, Nückles & Renkl, 2010; Hinds, 1999; Lentz & de Jong, 2006). This effect has also been found to apply to teachers (e.g. Nathan & Koedinger, 2000). Teachers, usually referred to as being domain experts in their content area, as they possess a high level of specialized knowledge, are considered to be prone to an expert blind spot (Nathan & Petrosino, 2003) when evaluating the difficulty of mathematical problems for students.

Following Nickerson's (1999) anchoring model, teachers may be inclined to use their specialized knowledge as an anchor when assessing the difficulty of problems for students. As a result, they are not able to take the student perspective adequately.

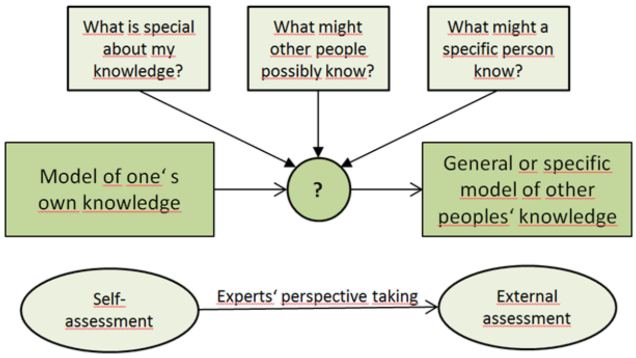

Figure 1: The process of perspective taking through anchoring and adjustment (adapted from Nickerson, 1999)

According to Nickerson (1999, see Figure 1), people tend to build an inaccurate mental model of the potential knowledge of general or specific others. They fail to take into account the specificity or exclusivity of their own knowledge, therefore unconsciously using it as an anchor when estimating other persons' knowledge. As a result, teachers might underestimate the difficulty that a problem will impose on a student, and overestimate students' performance.

However, the ability to adequately assess the difficulty of tasks for students is a crucial aspect of teaching expertise. It is necessary for communicating efficiently with students as well as for adapting teaching behaviour in and outside the classroom (e.g. selecting problems for homework, lessons or exams). Teachers should be able to take a novice's perspective and judge task attributes independently of their own perception of difficulty or effortlessness (Helmke, Hosenfeld & Schrader, 2004).

Cognitive Load Theory

Cognitive Load Theory (CLT; e.g. Sweller, 2005; Sweller, van Merrienboer & Paas, 1998) can help to understand how and why experts and novices differ in their perceptions of task difficulty and how one could deal with these discrepancies.

Working Memory Capacity And Perceived Task Difficulty According to CLT, every learning process is associated with cognitive mental load. The extent, to which a learner experiences this mental load, depends on the degree of the learner's expertise regarding the subject. Experienced learners use already existing knowledge structures, so called schemas. Schemas serve as patterns that help to structure and integrate incoming information (Sweller, 1994). But often, *new* information is being processed that needs the learner's full working memory capacity. If cognitive schemas do not exist and yet have to be built, working memory, the capacity of which is limited, is loaded to a high extent.

Perceived task difficulty according to CLT should mainly be a function both of the intrinsic mental load imposed on the learner (i.e., the complexity or difficulty of a task) and the amount of extraneous mental load (i.e., the load induced by an ineffective instructional design of the task). Generally, extraneous load has been shown to be an important factor hindering effective learning (e.g. Paas & van Merrienboer, 1994). Intrinsic load cannot be influenced, as it is inherent to the task itself and can only be moderated by the amount of a learner's prior knowledge. In contrast, extraneous load can and should be reduced. Once working memory is disburdened of extraneous load, more working memory capacity is available for understanding and schema acquisition.

Instructional Techniques Reducing Extraneous Mental Load Novice learners should be provided with learning material designed according to principles derived from CLT. The main principles are: *integrated-format* (Sweller, 2005), s*tep-by-step-guidance* (Kalyuga, Chandler & Sweller, 2001) or *worked examples* (Renkl, 2005). Tasks following these design principles substantially reduce the amount of extraneous load imposed on the learner.

An *integrated-format* in task design as compared to a *split-attention* design (Sweller, 2005) facilitates learning, as the learner does not have to search and integrate relevant information by himself, before passing on to the solution. With this procedure, information is presented close to each other and allows for an easier processing. A s*tep by step guidance* (e.g., Kalyuga, Chandler & Sweller, 2001) helps the learner to solve a problem without struggling to find all needed solution steps in a correct order. Instead, a processing guideline is given, leaving more working memory capacity available for the understanding of the single steps. *Worked examples* (e.g., Renkl, 2005), as compared to traditional problem solving techniques, consist of a problem, elaborated solution steps and the solution itself. Again, working memory capacity is free from extraneous load, as no potentially irrelevant trial and error processes are performed. This, once again, results in better schema acquisition and deeper understanding.

Expertise Reversal Effect It is important to emphasize that the effects of the just mentioned CLT principles are only prevailing with regard to novice learners. The positive learning outcome of material that is designed for novice learners may, in contrast, be reversed for experts. The guidance or additional information given by *optimized* learning material (from now on, the term *optimized* will be used with regard to learning tasks that are designed in favour of *novice* learners) can interfere with experts' advanced cognitive structures and schemas that have already been built. Kalyuga (2007) named this phenomenon *expertise reversal effect*. He described that an optimized learning tasks is experienced as being more difficult to process and causes a redundancy effect, when presented to expert learners. This results in *increased* extraneous load and worse performance.

From this follows that the same learning material may cause *reversed* effects for novice and experienced learners. However, as experts may perceive optimized tasks as being more difficult than non-optimized tasks, they may also be subject to an expert blind spot when assessing the potential difficulty of the tasks for novice learners. This prediction is in line with Nickerson's anchoring and adjustment model (1999). Experts judge optimized learning material as being difficult to solve, use that judgement as an anchor for estimating novices' performance and thus underestimate novices' performance on these tasks. The opposite is true for non-optimized items, resulting in an overestimation of novices' performance.

Teachers as domain experts *and* educators should be knowledgeable of this expertise reversal effect and able to estimate the difficulty of tasks for students as novice learners *independently* of their own experienced mental load. In the present study we investigated whether this assumption is true for two groups of mathematics experts.

Research Questions and Predictions

In the present study, we investigated whether pre-service as well as in-service mathematics teachers are subject to an expert blind spot when judging the difficulty of problems for students and whether the two expert groups differ in their estimations. Differences in estimations can be expected due to different levels of teaching experience. The tasks presented to the expert groups varied in instructional design according to CLT, but were comparable in complexity, thus keeping intrinsic cognitive load stable.

Following our theoretical assumptions, both pre-service and in-service teachers should use their expert knowledge as an anchor and underestimate the difficulty of the tasks for novice students in general.

1) Therefore, we predicted that both expert groups would generally overestimate the amount of tasks that novice students would be able to solve correctly (*overestimation hypothesis*).

An anchoring effect should manifest itself in highly correlated ratings of one's own perceived mental load and estimated task performance of novice learners.

2) Hence, we predicted that the correlation between the experts' self-rated mental load and estimated task performance of novice learners is significantly larger than the correlation between estimated student performance and students' actual performance (*anchoring hypothesis*).

Following expertise reversal effect, we further expected that the experts would experience less mental load when solving non-optimized than optimized tasks.

3) Consequently, the overestimation of novices' performance should be significantly larger with regard to non-optimized tasks as compared with optimized tasks (*expertise reversal hypothesis*).

Method

Participants

Thirty-four pre-service teachers majoring in mathematics (mean study time being 6.12 semesters, *SD* = 3.18) and 13 in-service mathematics teachers (mean time of working experience being 12.85 years, *SD* = 9.13) participated in the study. Two different expert groups were chosen to allow for possible conclusions regarding experience levels. Whereas pre-service teachers usually do not have school teaching experience, the amount of in-service teachers' teaching experience could become evident in their ratings of students performance.

Both expert groups' estimations were compared to solution frequencies collected from 54 9^{th} grade high school students (mean age being 14.26, *SD* = .52). All participants attended the study on a voluntary basis and received financial compensation.

Study design

We used level of expertise (pre-service teachers vs. in-service teachers vs. novices) and the instructional design of the task (non-optimized vs. optimized mathematical problems) as independent variables. Dependent variables encompassed experts' perceived mental load and performance, their estimations of novices' performance and novices' actual performance on a number of mathematical tasks. Estimations were compared to students' actual performance

Instrument and measures

Two mathematics experts created ten tasks on algebra, geometry and trigonometry. To achieve a high level of curriculum validity, contents of the tasks were chosen to meet the requirements expected from pupils on that 9^{th} grade school level (e.g. calculation of area, theorem of Pythagoras, angular sum). Each task was designed in a non-optimized and optimized version. Tasks without didactic optimization were adapted from mathematics problems currently used in school. Tasks optimization was achieved by using one of the following CLT design principles (the latter being the optimized design): either *split-attention-format* vs. *integrated-format;* or *traditional problem solving* vs. *step-by-step-guidance;* or *traditional problem solving* vs. *worked examples*. A task on angular sum, for example, was either designed with help of a diagram and angular degrees being spread over the working sheet making it difficult to match needed information, or presented with a diagram and angular degrees being close to each other (optimized; integrated format). So, whereas each task covered exactly the same mathematical problem (keeping intrinsic cognitive load stable), the design of the task (extraneous cognitive load) varied, allowing for the measurement of differences in mental load and performance due to task design.

Perceived mental load was assessed by the following question adapted from cognitive load literature: "*How difficult did you find working on the task?*", and measured on a six-point rating-scale ranging from "*not at all difficult*" to "*very difficult*".

Teachers' estimations of student performance were collected by using a prototype description of a fictional 9^{th} grade high school class: "*Imagine that you are the teacher of a class with 30 students, all having different achievement levels; there are very good, average and very poor students. Now, you want to use the same task that you have just worked on for an exam. How many students of this class will presumably solve the task correctly?"*

Participants' task performance was measured as the number of correctly solved tasks (the maximum score being ten). Each of the participants' solutions was rated by two independent mathematics experts as being correctly or falsely solved. When no accordance could be initially found, the two experts discussed their different ratings and agreed on one in a second step.

Procedure

Each participant received a booklet with ten tasks. Five of the tasks were presented in a non-optimized version and five were presented in an optimized version, balanced within the booklet. Furthermore, each task presented in its non-optimized version (e.g. angular sum, *split-attention-format*) had a corresponding item in its optimized version (e.g. analogous angular sum task, *integrated-format*), placed elsewhere in the booklet. Using this method, repetition effects by having the participants solving the same task twice were avoided, but still estimations based on both task designs were collected.

The participants solved each mathematical problem within a fixed period of time. The time constraint should prevent ceiling effects from occurring. Experts, given unlimited time to solve the tasks, perceive only little to no mental load, as enough working memory capacity is free for solving most tasks correctly, no matter which design is presented. Under these circumstances, an effect of task design on experienced mental load can no longer be detected (Paas, Renkl & Sweller, 2003).

After having solved each task, participants rated their perceived mental load on a six-point rating scale. Then, they estimated how many students of the fictional 9^{th}-grade class would be able to solve the tasks they have just worked on correctly. After having finished the rating process, participants continued with the next mathematical problem. At the end, demographic data was collected.

Results

In a first step, we compared pre-service teachers' and in-service teachers' perceived mental load and performance for both item types. For this purpose, each participant's ratings and performance data was aggregated (five for non-optimized and five for optimized tasks) and then compared with a paired t-test.

In line with CLT, pre-service teachers experienced significantly more mental load when solving optimized ($M = 2.33$, $SD = 0.67$) than non-optimized tasks ($M = 2.11$, $SD = 0.58$), $t(33) = -2.08$, $p < .05$. However, pre-service teachers did not significantly perform worse on optimized ($M = 75.88\%$, $SD = 20.17\%$) than on non-optimized tasks ($M = 78.82\%$ $SD = 16.29\%$), $t(33) = 0.82$, *ns*.

In-service teachers experienced no significantly different degree of mental load for optimized ($M = 2.72$, $SD = 0.72$) and non-optimized tasks ($M = 2.59$, $SD = 0.94$), $t(12) = -1.13$, *ns*. Also, performance for optimized ($M = 76.92\%$, $SD = 17.97\%$) and non-optimized tasks ($M = 73.85\%$, $SD = 18.94\%$) did not differ significantly, $t(12) = -.56$, *ns*.

In a second step, students' solution frequencies were analysed. In line with CLT, the 9^{th}-grade students solved more optimized ($M = 51.11\%$, $SD = 26.51\%$) than non-optimized tasks ($M = 44.07\%$, $SD = 24.69\%$), $t(53) = 2.38$, $p < .05$ (all performance data are presented in Figure 2).

Finally, performance data between the participant groups were compared in a repeated measures ANOVA. Both pre-service ($F(1,86) = 44.08$, $p<.01$) and in-service teachers ($F(1,65) = 16.69$, $p<.01$) solved significantly more tasks than students did, whereas performance between the expert groups ($F(1,45) = .16$, *ns*) did not differ significantly.

Overestimation Hypothesis

To test the overestimation hypothesis, we computed difference scores. Students' real solution frequencies for each item were subtracted from pre-service and in-service teachers' estimations of how many students would be able to solve this corresponding task correctly. A positive difference score thus indicated an overestimation and a negative score indicated an underestimation. Each participant's difference scores were then aggregated for item type (five scores for non-optimized and five for optimized tasks) and used for further analysis.

As predicted, pre-service teachers overestimated students' performance both on non-optimized tasks, $t\ (33) = 6.29$, $p < .01$, and optimized tasks, $t\ (33) = 2.34$, $p < .05$ (one-sample t-test). However, in-service teachers' general overestimation of student performance did not reach statistical significance both for non-optimized ($t\ (12) = 1.21$, *ns*) and optimized tasks ($t\ (12) = -.13$, *ns*). Overestimation scores between the expert groups did not differ significantly, $F(1,45)=3.19$, *ns.* (estimation data for pre-service and in-service teachers are presented in Figure 2).

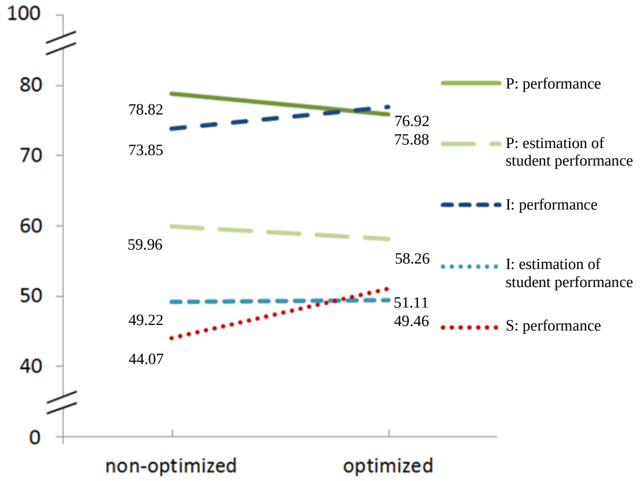

Figure 2: Pre-service teachers' (P) and in-service teachers (I) performance and estimation of student performance and students' (S) performance as function of task design (%)

Anchoring Hypothesis

To test for an anchoring effect, we computed and compared Fisher z transformed individual correlations. Pre-service and in-service teachers' mental load ratings for each item were correlated with their estimation of student performance for that particular item. Further, the estimation of student performance for each item was correlated with students' actual performance on that item. This procedure allowed analysing whether experts' estimations were closer to their perceived mental load or to students' actual performance. Experts' perceived mental load (as compared to experts' actual performance on each task) was used for the analysis. Whereas performance on a task cannot be determined immediately by the participants (as it remains unclear whether they solved a task correctly or not), mental load served as adequate and approximate measure of task difficulty. The individual correlations were aggregated (five correlations for non-optimized and five for optimized tasks) and then compared in a repeated measures ANOVA.

As predicted, results showed a significant difference between both correlation types, thus indicating an anchoring effect. Pre-service teachers' estimations of students' performance were significantly more strongly correlated with own perceived mental load than with students' actual performance, $F(1,33) = 169.45$, $p<.01$. A very similar pattern was found for in-service teachers' correlations, $F(1,12) = 35.64$, $p<.01$ (correlation coefficients for both expert groups are depicted in Figure 3). It can be concluded that both expert groups used their own perceived mental load as an anchor to estimate the difficulty that the tasks would impose on the students.

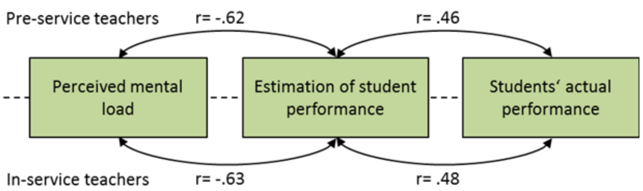

Figure 3: The anchoring effect in both expert groups

Expertise Reversal Hypothesis

As predicted by the expertise reversal hypothesis, both expert groups' overestimations were moderated by the instructional design of the tasks. Pre-service teachers' overestimation of student performance was significantly larger for non-optimized (*M* = 15.88%, *SD* = 14.74%) than optimized tasks (*M* = 7.15%, *SD* = 17.81%), t (33) = 4.21; $p < .01$. Also, in-service teachers' overestimation was significantly larger for non-optimized (*M* = 5.63%, *SD* = 16.82%) than optimized items (*M* = -.62%, *SD* = 16.77%), t (12) = 2.54; $p < .05$ (see Figure 2 for mean scores). Both expert groups seem to have failed to take into account the benefits of didactic optimization of the learning material for novice learners. As was already described in the "Overestimation Hypothesis" section, differences in overestimation scores did not to reach statistical significance for both expert groups.

Discussion

In the present study, we investigated whether pre-service and in-service mathematics teachers are subject to an expert blind spot, when judging the difficulty of tasks for students. The tasks were designed in accordance with didactic principles derived from CLT, which have differential effects on the learning outcome of experts and novices. Whereas novice learners experience a relief from extraneous mental load when being presented optimized learning material, thus having more working memory capacity available for schema acquisition and therefore performing better on those tasks, the opposite is true for expert learners. These learners, being presented with optimized tasks, experience increased extraneous mental load (due to a redundancy effect) and judge those tasks not only as being more difficult to work on for themselves, but also as being more difficult to solve for novice learners. The reason for this misjudgement lies in an anchoring effect, as experts generally use their own knowledge base and estimations as ground for judging the difficulties that other persons (in this case: novices) may have. To test these assumptions, experts' mental load ratings while working on mathematical tasks and their estimations of novice performance were compared to real solution frequencies obtained from novices.

Results indicate an egocentric bias, as the experts' general estimations for student performance were highly correlated with their own experienced mental load. Especially, the overestimation of students' task performance was significantly larger for non-optimized than optimized items, indicating an expert blind spot that can be interpreted in terms of an expertise reversal effect. Experts failed to adequately take into account the beneficial or detrimental effects of didactical variation in task design. Rather, they judged both non-optimized and optimized mathematical tasks as being equally difficult for students, which in fact was not the case in our student sample.

However, only pre-service teachers' general overestimation of student performance was significant, whereas in-service teachers' overestimation failed to reach statistical significance. Relating to Nickerson's (1999) anchoring and adjustment model, in-service teachers seem to have a more accurate mental model of students' knowledge than pre-service teachers do. Teaching experience seems to have had a debiasing effect on an egocentric bias, thus resulting in better judgements of student performance.

Nevertheless, there are certain limitations to the present study. The first one concerns the yet small sample size of in-service teachers compared to pre-service teachers. The results obtained so far should be further consolidated by equalizing sample sizes for both expert groups, thus allowing for a better comparability and generalizability. This would allow for a detailed analysis of the variability in teaching experience between in-service teachers and its effects on the estimations of student performance. Also, though in-service teachers showed no different estimation pattern for both item types than pre-service teachers do, the overall level of overestimation was a different one. With a bigger sample size, this issue could be further investigated and possible influencing variables could be detected.

Another limitation concerns the actual level of expertise in both teacher groups. As presented in the results section, pre-service and in-service teachers solved significantly more tasks correctly that students did. This allows the conclusion that both teacher groups have more specialized knowledge as compared to students and can indeed be called experts. Also, it is not necessarily surprising that pre-service teachers, not yet having gained teaching experience and being presented with learning tasks obtained from school books, do not solve the mathematical tasks in large part. However, it remains unclear why in-service teachers with a high level of specialized knowledge as well as teaching experience only show similar performance rates on the mathematical problems instead of solving almost all of them correctly.

Finally, the present study does not allow for a detailed insight into participants' estimation processes. After having rated each mathematical task, participants had the opportunity to answer an open-format question and give additional information on what they thought made each tasks difficult or easy. This possibility was barely used, thus not allowing for any further insights into participants' cognitive processes while judging the difficulty of the tasks for students.

Future research will address the just mentioned issues and explore ways in which teachers' ability to see the difficulty of tasks from a student's perspective can be improved and be emphasized already in teacher education. To the authors' knowledge, no research so far has examined experts estimations of novices' performance using instructional design principles derived from CLT. Subsequent studies with different participant groups shall shed more light on anchoring and adjustment processes in experts. Experts' cognitive processes while solving the mathematical tasks shall be further investigated. Also, longitudinal designs could be conducted in order to analyse effects of intervention programs on teachers' perception of learning material.

References

Helmke, A., Hosenfeld, I., & Schrader, F.-W. (2004). Vergleichsarbeiten als Instrument zur Verbesserung der Diagnosekompetenz von Lehrkräften. In R. Arnold & C. Griese (Eds.), *Schulleitung und Schulentwicklung* (pp. 119-144). Hohengehren: Schneider.

Herppich, S., Wittwer, J., Nückles, M., & Renkl, A. (2010). Do tutors' content knowledge and beliefs about learning influence their assessment of tutees' understanding? In S. Ohlsson & R. Catrambone (Eds.), *Proceedings of the 32th Annual Conference of the Cognitive Science Society* (pp. 314-319). New York: Erlbaum.

Hinds, P. (1999). The curse of expertise: The effects of expertise and debiasing methods on predictions of novice performance. *Journal of Experimental Psychology: Applied, 5,* 205-221.

Kalyuga, S. (2007). Expertise reversal effect and its implications for learner-tailored instruction. *Educational Psychology Review, 19*, 509-539.

Kalyuga, S., Chandler, P., & Sweller, J. (2001). Learner experience and efficiency of instructional guidance. *Educational Psychology, 21*(1), 5-23.

Lentz, L., & De Jong, M. (2006). The curse of expertise: Exploring the problem of anticipating readers' needs. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?rep=rep1&type=pdf&doi=10.1.1.136.1732

Nathan, M., & Koedinger, K. (2000). Teachers' and researchers' beliefs about the development of algebraic reasoning. *Journal for Research in Mathematics Education, 31*(2), 168- 190.

Nathan, M., & Petrosino, A. (2003). Expert blind spot among preservice teachers. *American Educational Research Journal, 40*(4), 905-928.

Nickerson, R. S. (1999). How we know - and sometimes misjudge - what others know: Imputing one's own knowledge to others. *Psychological Bulletin, 125,* 737-759.

Paas, F., & van Merrienboer, J. (1994). Variability of worked examples and transfer of geometrical problem-solving skills: A cognitive-load approach. *Journal of Educational Psychology, 86*(1), 122-133.

Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive load theory and instructional design: Recent developments. *Educational Psychologist, 38*(1), 1-4.

Renkl, A. (2005). The worked-out-example principle in multimedia learning. In R. E. Mayer (Ed.), *Cambridge Handbook of Multimedia Learning* (pp. 229-246). Cambridge: University Press.

Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction, 4*, 295-312.

Sweller, J. (2005). Implications of cognitive load theory for multimedia learning. In R. E. Mayer (Ed.), *Cambridge handbook of multimedia learning* (pp. 19–30). New York: Cambridge University Press.

Sweller, J., van Merrienboer, J., & Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10* (3), 251-196.

Tversky, A., & Kahneman, D. (1974). Judgement under uncertainty: Heuristics and biases. *Science, 185,* 1124-1131.