Benefits for Processes Cause Decrements in Outcomes: Training Improves Tutors' Interactivity at the Expense of Assessment Accuracy

Stephanie Herppich (stephanie.herppich@sowi.uni-goettingen.de)
University of Göttingen, Educational Institute
Waldweg 26, 37073 Göttingen, Germany

Jörg Wittwer (joerg.wittwer@sowi.uni-goettingen.de)
University of Göttingen, Educational Institute
Waldweg 26, 37073 Göttingen, Germany

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

Alexander Renkl (renkl@psychologie.uni-freiburg.de)
University of Freiburg, Department of Psychology, Developmental and Educational Psychology
Engelbergerstrasse 41, 79085 Freiburg, Germany

Abstract

Tutoring gives tutors the opportunity to engage in interactive strategies that help them to assess a tutee's understanding. However, tutors without teaching experience often do not engage in interactive strategies and, thus, have difficulty with accurately assessing a tutee's understanding. We conducted an experiment with 39 tutor-tutee dyads to test whether tutors who received training in interactive strategies would become more interactive and more accurate in assessing a tutee's understanding. Results showed that trained tutors provided a more interactive style of tutoring than untrained tutors. However, due to being more interactive, trained tutors produced less accurate assessments than untrained tutors. This suggests that changing the style of tutoring to implement interactive strategies puts a high burden on a tutor's cognitive capacity. Hence, there is obviously little cognitive capacity left that could be used to assess a tutee's understanding. Training methods that automate strategy use might enhance a tutor's assessment accuracy.

Keywords: one-on-one human tutoring; training; tutoring interactions; assessment accuracy

Introduction

In one-on-one tutoring, tutors have the possibility to engage in interactive tutoring strategies such as asking questions or providing hints. When a tutee responds to a tutor's interactive tutoring strategies, for example, by answering a question, a tutor can learn what a tutee does and does not know (Chi, 2009; Hmelo-Silver & Barrows, 2006). Thus, in the course of tutoring, a tutor has the opportunity to collect a multitude of information that can be used to summatively assess a tutee's understanding after tutoring session. This summative assessment may also help a tutor to prepare the next tutoring session by choosing material that is suited to a tutee's individual level of understanding (e.g., Chi, Jeong, & Siler, 2004; Kalyuga, 2007; cf. also the discussion of the concept of interim assessments for the school context by Perie, Marion, & Gong, 2009).

However, research has shown that inexperienced tutors, that is, tutors who are not trained in teaching (Chi et al., 2001; Graesser, D'Mello, & Cade, 2011), often do not engage in interactive tutoring strategies. Instead, they frequently dominate tutoring by providing lengthy explanations (e.g., Chi et al., 2001; Cromley & Azevedo, 2005). In addition, inexperienced tutors regularly fail to assess a tutee's understanding accurately (Chi et al., 2004; Herppich et al., 2013b).

Against this background, we conducted an experimental study to test whether inexperienced tutors who received training in interactive tutoring strategies would be able to implement an interactive style of tutoring. We were interested in whether a more interactive style of tutoring would benefit a tutor's assessment of a tutee's understanding after tutoring.

Tutoring Strategies of Experienced and Inexperienced Tutors and Their Influence on Assessment

In contrast to inexperienced tutors, experienced tutors are trained or experienced in teaching (cf. Cromley & Azevedo, 2005; D'Mello et al., 2010; McArthur, Stasz, & Zmuidzinas, 1990). Research shows that experienced tutors tend to provide a different style of tutoring than do inexperienced tutors. More specifically, experienced tutors more often engage in interactive tutoring strategies than inexperienced tutors. For example, they frequently scaffold a tutee by providing hints or asking questions (Cade et al., 2008; Chi, Roy, & Hausmann, 2008; Cromley & Azevedo, 2005). Scaffolding is a genuinely interactive tutoring strategy because it elicits constructive responses from a tutee (Hmelo-Silver & Barrows, 2006). In this vein, Herppich et al. (2013a, 2013b) found that experienced tutors caused tutees to utter more knowledge deficits, that is, incomplete beliefs, incorrect beliefs, or misconceptions, in the course of tutoring than inexperienced tutors. In addition, experienced tutors were more accurate in assessing a tutee's understanding after tutoring than inexperienced tutors. The results suggest that a tutee's uttered knowledge deficits are diagnostically informative because they indicate what a tutee does not know (cf. Chi, et al., 2004; Cromley & Azevedo, 2005). Thus, tutors might derive information from these knowledge deficits that can be used to assess a tutee's understanding after tutoring.

Training Inexperienced Tutors

To test whether training inexperienced tutors in interactive tutoring strategies would improve their style of tutoring, we developed a training method that aimed at prompting inexperienced tutors to abstain from giving lengthy explanations and, instead, to engage in more interactive tutoring strategies such as scaffolding (cf. Chi, et al., 2008). As a result of implementing more interactive tutoring strategies in the course of tutoring, tutors were assumed to more intensively engage in collecting diagnostically relevant information that could be used to assess a tutee's understanding after tutoring.

Based on what is known about effective training methods in the domain of learning strategies (Mandl & Friedrich, 1992), the development of our training method was guided by several principles. First, training methods should inform about the advantages associated with the strategies targeted in the training. Second, training methods should directly convey knowledge about the strategies that need to be trained. Third, training methods should help to practice the targeted strategies (Klauer, 1988; Mandl & Friedrich, 1992). Research has shown that training methods that are in accordance with these principles are particularly effective (Dignath, Buettner, & Langfeldt, 2008; Leutner, Leopold, & Elzen-Rump, 2007).

By now, little attention has been given to training methods that aim at fostering an interactive tutoring style in the service of improving assessment accuracy. However, existing research on training tutors with the aim of enhancing a tutee's learning has well documented that tutors are often able to spontaneously implement the strategies that are targeted in training. Yet, tutors have difficulty with changing their style of tutoring in the long run (King, Staffieri, & Adelgais, 1998). Moreover, even though tutors are able to change their tutoring strategies, this might not necessarily increase the effectiveness of tutoring (Chi et al., 2001). In their review on tutoring-based instruction, Graesser et al. (2011) summarized research on tutor training in the following way:

> …it is difficult to train tutors to adopt particular strategies. They rely on their normal conversational and pedagogical styles.… it is difficult to force the human tutors to adopt changes in their language and discourse, particularly those levels that are unconscious and involuntary. (p. 422).

Hypotheses

In this study, we tested the effectiveness of a training method that aimed at helping tutors to implement a more interactive style of tutoring. We addressed the following hypotheses:

1) Trained tutors engage in more interactive tutoring strategies in the course of tutoring than untrained tutors.
2) Trained tutors are more accurate in assessing a tutee's understanding after tutoring than untrained tutors.
3) The more interactive style of tutoring explains why trained tutors are more accurate than untrained tutors in assessing a tutee's understanding after tutoring.

Method

Sample and Design

A total of N = 39 dyads of tutors and tutees participated in the experiment. The topic of tutoring was the human circulatory system. All tutors were university students majoring in biology with a mean age of 22.38 years (SD = 2.47). Thirty-five tutors were female and 4 tutors were male. Twenty tutors received training in interactive tutoring strategies (= *trained tutors*), whereas 19 tutors received no training (= *untrained tutors*). As indicated by a multiple-choice test, all tutors had sufficient knowledge about the human circulatory system. There was no significant difference in knowledge between trained tutors (M = 8.45, SD = 2.26) and untrained tutors (M = 8.26, SD = 1.78), $F(1, 37) = 0.81$, $p > .05$, $\eta^2 < .01$ (small effect). Moreover, trained (mean rank = 18.88) and untrained tutors (mean rank = 21.18) did not differ in their previous experience in providing tutoring, coded as 1 = *no experience*, 2 = *sporadic tutoring*, 3 = *regular tutoring*, U = 167.50, z = -0.69, $p > .05$, r = -.11 (small effect). Tutees were seventh-grade students from the middle track of the German school system (i.e., from *Realschulen*). Of the tutees, 9 were female and 29 were male; one tutee did not indicate gender.

Tutors were randomly assigned to the two experimental conditions (training vs. no training) and tutees were randomly assigned to tutors. The dependent variables in this experiment were the extent to which a tutor elicited knowledge deficits from a tutee in the course of tutoring and the accuracy with which a tutor assessed a tutee's understanding after tutoring.

Materials

Textbook Passage (Tutees and Tutors) In the tutoring session, the tutor-tutee dyads engaged in a dialogue based on a passage about the human circulatory system. We adapted this passage from the study by Chi et al. (2001). The passage consisted of 59 sentences and each sentence was printed on a separate sheet of paper. The sentences were presented to the tutor and the tutee in a ring binder.

Concepts Test (Tutees and Tutors) We used a shortened version of a test that was employed by Herppich et al. (2013b). This shortened version consisted of 16 multiple-choice items that assessed a tutee's understanding of concepts about the human circulatory system. For example, it included the following item: What is the task of the heart in the human organism? (1) The heart pumps the blood. (2) The heart cleans and filters the blood. (3) The heart supplies the blood with oxygen. (4) Don't know. The items of the original test were adapted from tests developed by Sungur and Tekkaya (2003) and by Michael et al. (2002) or constructed on the basis of the literature on misconceptions of the human circulatory system (e.g., Pelaez et al., 2005). A correct answer indicated a scientifically correct understanding of the concept. Each of the incorrect answers indicated a specific type of incorrect understanding of the concept. Hence, a tutee could achieve a maximum number of 16 points in the concepts test.

To examine the accuracy with which the tutors assessed a tutee's understanding of the human circulatory system after tutoring the tutors were also administered the test.

Training in Interactive Tutoring Strategies (Trained Tutors) The trained tutors received training in interactive tutoring strategies. The training took about 45 minutes and was presented on a computer screen. The training aimed at helping the trained tutors to adopt interactive tutoring strategies that would enable them to elicit knowledge deficits from a tutee. The training consisted of two building blocks. In the first building block, the trained tutors were informed about the problem that tutors often are not interactive and, thus, cannot accurately assess a tutee's understanding (Brown, Campione, & Day, 1981). Subsequently, the trained tutors were provided with information about three strategies, namely, (1) abstaining from giving lengthy explanations, (2) intensifying question asking, and (3) increasing scaffolding in response to a tutee's contribution (Cade et al., 2008, Chi et al., 2008; Herppich et al., 2013a). To learn about the three strategies, the trained tutors first read an explanatory text and then watched two videos of fictitious tutoring sessions. The first video presented a tutor who failed to engage in interactive tutoring strategies and, thus, to receive information about a tutee's understanding. The second video, in contrast, presented the same tutor who did engage in interactive tutoring strategies, which helped the tutor to receive information about a tutee's understanding (cf. Renkl, 2005). In the second building block, trained the tutors also watched videos that presented positive and negative examples of tutoring strategies. This time, however, the tutoring strategies were not explained to the trained tutors. Instead, the trained tutors were prompted to self-explain what constituted the difference between the positive and negative examples. More specifically, the trained tutors were asked to provide information about the tutoring strategies that they saw in the videos and about the effects of such tutoring strategies for assessing a tutee's understanding (cf. Renkl, 2005). Finally, the trained tutors were required to indicate what they would do in order to change the tutoring strategies that they saw in a negative example. This was done to actively stimulate the application of the to-be-learned strategies (cf. Klauer, 1988).

Introductory Text (Untrained Tutors) Instead of receiving training in interactive tutoring strategies, the untrained tutors read a short text. The text provided information about the effectiveness of tutoring and about problems associated with assessing a tutee's understanding. However, the untrained tutors did not receive any instruction on how to solve these problems. Instead, they were asked to provide tutoring in whatever manner they assumed appropriate.

Procedure

Each tutoring session was divided into three phases: pretest phase, tutoring phase, and posttest phase. On average, a tutoring session lasted about 3 hours.

In the pretest phase, each tutee and each tutor individually read the passage about the human circulatory system. Afterwards, the trained tutors received training and the untrained tutors read the text.

In the tutoring phase, tutor-tutee dyads jointly read the passage about the human circulatory system sentence-by-sentence and engaged in a dialogue about each sentence. All tutoring phases were videotaped.

In the posttest phase, the tutees completed the concepts test. The tutors also received the items of the concepts test and were asked to indicate for each item which of the given response options the tutee would choose.

Codings and Analyses

Elicitation of Knowledge Deficits (Tutors) As an indicator of engaging in interactive tutoring strategies, we coded the knowledge deficits that a tutor elicited from a tutee. To do so, we used a coding scheme adapted from Chi et al. (2004). Every knowledge deficit that a tutee uttered was coded from its beginning to its end (event sampling procedure).

We coded a knowledge deficit whenever a tutor elicited from a tutee an utterance that (1) contradicted a piece of knowledge stated in the textbook passage, that (2) was incomplete, that (3) was vague, that (4) was incorrect and not addressed by the textbook passage, or when the tutee (5) did not utter a certain piece of information at all, that is, the tutee obviously missed this piece of information. In one tutoring session, for example, the tutor asked: "Why does the blood need to go to the lung? What does the lung do?" And the tutee answered: "Yes, um, yes, the lung filters the blood." This answer was coded as utterance of a knowledge deficit because it represents a normatively incorrect understanding. To standardize coding, the coder used a written instruction. For each tutor-tutee dyad, we summed up the number of elicited knowledge deficits.

Summative Assessment (Tutors) To examine the accuracy with which a tutor assessed a tutee's understanding of the human circulatory system after tutoring, we compared a tutee's responses in the concepts test with a tutor's estimations of a tutee's responses in the concepts test. To do so, we made the comparison on an item-by-item basis (cf. Hoge & Coladarci, 1989). Hence, a tutor could achieve a maximum score of 16 points. Higher scores indicated a higher assessment accuracy.

Mediation Analysis To test our hypotheses, we performed a mediation analysis. We calculated total, direct, and indirect effects in accordance with our hypotheses by applying regression-based path analysis. To test for the statistical significance of an indirect effect, we derived 95% confidence intervals for indirect effects as well as standard errors for indirect effects via bias-corrected bootstrap (for guidelines, see, e.g., Hayes, 2009, 2012). This approach resolves some methodological problems associated with the Sobel test (Hayes, 2009).

Results

For all analyses, we used an alpha level of .05. For directional hypotheses, we used one-tailed tests. In the analyses, trained tutors were coded as 1 and untrained tutors were coded as 0. As effect size for indirect effects in the mediation analysis, we report κ^2. According to Preacher and Kelley (2011), effects are small when $\kappa^2 = .01$, medium when $\kappa^2 = .09$, and large when $\kappa^2 = .25$. All analyses were performed using SPSS 20.0.0, the PROCESS macro for SPSS introduced in Hayes (2012; to perform the mediation analysis), and AMOS 20.0.0 (to receive standardized path coefficients for the mediation analysis). Table 1 shows the means and standard deviations of the dependent variables.

Table 1: Means and standard deviations (in parentheses) of the experiment's dependent variables

Variable	Trained Tutors *M* (*SD*)	Untrained Tutors *M* (*SD*)	All Tutors *M* (*SD*)
Elicited Knowledge Deficits	71.30 (40.46)	32.11 (28.63)	52.21 (40.01)
Assessment Accuracy	8.05 (2.54)	8.21 (2.30)	8.13 (2.40)

Impact of Training on Implementing Interactive Tutoring Strategies

Our first hypothesis stated that trained tutors would more often engage in interactive tutoring strategies than untrained tutors. Thus, trained tutors should elicit more knowledge deficits from their tutees than untrained tutors. As can be seen in Figure 1, trained tutors elicited more utterances of knowledge deficits from their tutee than did untrained tutors, $R^2 = .25$, $F(1, 37) = 12.08$, $p < .05$, 95% CI [.26, .74]. Hence, the trained tutors in fact engaged in more interactive tutoring strategies than the untrained tutors.

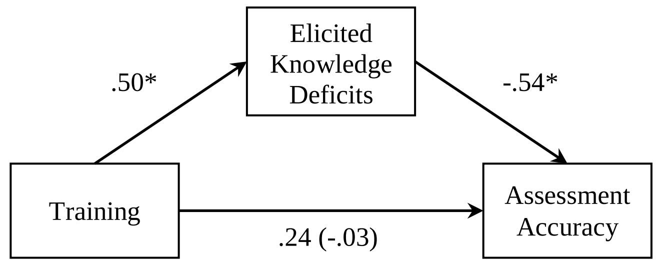

Figure 1: Mediation model for the effect of tutor training on assessment accuracy explained by the number of expressed knowledge deficits a tutor elicited from a tutee. Numbers represent standardized path coefficients for direct effects and, in parentheses, the total effect of the independent variable on the dependent variable. $^*p < .05$.

Impact of Training on Summative Assessment

Our second hypothesis stated that trained tutors would more accurately assess a tutee's understanding after tutoring than untrained tutors. However, as the total effect depicted in Figure 1 shows, there was no significant difference in assessment accuracy between trained tutors and untrained tutors, $R^2 < .01$, $F(1, 37) = 0.04$, $p > .05$, 95% CI [-.31, .24]. Hence, if only zero-order relations are taken into account, training tutors to implement interactive tutoring strategies failed to exert an influence on assessment accuracy.

Interactive Tutoring Strategies as Mediator

Our third hypothesis stated that the higher number of a tutee's elicited knowledge deficits would explain why trained tutors assessed a tutee's understanding after tutoring more accurately than untrained tutors. To statistically test this hypothesis, we computed the indirect effect even though the total effect (i.e., the effect of training on assessment accuracy) was not significant (cf. Hayes, 2009; Shrout & Bolger, 2002). To test the indirect effect, we constructed a bias corrected 95% bootstrap confidence interval as well as bootstrap standard errors from 10000 bootstrap samples. We found a significant negative indirect effect indicating that implementing interactive tutoring strategies as a result of receiving training decreased assessment accuracy with a standardized point estimate of -.27 ($SE = .10$), 95% CI [-.46, -.12], $\kappa^2 = .26$ (zero-order correlation between elicited knowledge deficits and assessment accuracy: $r = -.43$, $p < .05$). Translated to unstandardized estimates, the number of items correctly estimated by trained tutors was 1.28 points ($SE = 0.54$) lower (and not higher) than the number of items correctly estimated by untrained tutors as mediated by the number of elicited knowledge deficits.

Discussion

This study examined the effectiveness of a training method that aimed at helping tutors to engage in interactive tutoring strategies in the course of tutoring. It was assumed that engaging in interactive tutoring strategies would benefit a tutor's assessment of a tutee's understanding after tutoring.

First, we found that trained tutors in fact showed a more interactive style of tutoring than untrained tutors. Hence, even though the duration of our training was rather short, it was obviously sufficient to help the tutors to implement more interactive tutoring strategies. As a result, tutees tutored by trained tutors more often uttered knowledge deficits than tutees tutored by untrained tutors. This finding is consistent with the results obtained by Herppich et al. (2013a).

Second, however, the trained tutors failed to assess a tutee's understanding more accurately than the untrained tutors. The trained tutors were even less accurate than the untrained tutors. As show by the mediation analysis, this result was explained by the greater extent to which trained tutors engaged in interactive tutoring strategies as a result of receiving training. This effect was probably not observable in the zero-order analysis because the two paths making up the indirect effect were opposite in sign (cf. Hayes, 2009).

An explanation for why trained tutors and untrained tutors did not differ in assessment accuracy, as indicated by the total effect in the mediation analysis, is that the changes in the tutoring strategies due to receiving training might not have been sufficient to produce changes in assessment accuracy. This explanation would be in accordance with the results obtained by Roscoe and Chi (2007), who found that strategies of tutors can only be influenced to a certain extent. Hence, in the context of the present study, the information gained from being more interactive might not have been enough to generate more accurate assessments (cf. Graesser et al., 2011).

However, it still remains an open question as to why the elicitation of knowledge deficits was detrimental for assessing a tutee's understanding after tutoring, as indicated by the indirect effect in the statistical analysis. First, it might be that trained tutors and untrained tutors differed in the types of knowledge deficits they elicited from a tutee. Eliciting a larger number of scientifically incorrect utterances as compared to missing knowledge pieces, for example, might have been more informative for the summative assessment. This is because the incorrect response options in the concepts test were based on common types of incorrect understanding of a concept (e.g., Pelaez et al., 2005). However, the relative number of knowledge deficits elicited per category did not differ significantly between trained tutors and untrained tutors for any of the five categories of knowledge deficits coded.

Second, the detrimental effect of eliciting knowledge deficits on summative assessment might be related to our measure of summative assessment accuracy. During the training, the tutors were repeatedly informed that a tutor should *get a picture of a tutee's understanding*. As a consequence, the trained tutors might have conceived a tutee's understanding on a more global level than on the level of conceptual understanding. Thus, after having completed the training, being more interactive and receiving more information from the tutees could have drawn the tutors' attention away from the knowledge they were to assess in the concepts test. This conjecture could be tested in future research that uses measures of assessment accuracy that are as manageable for tutors as a multiple-choice test on conceptual knowledge but that would tap different levels of a tutee's understanding.

Third, another explanation refers to the fact that the tutors in this study did not possess teaching experience. Hence, the interactive tutoring strategies targeted in the training might have been quite unfamiliar to the tutors. As a result, implementing interactive tutoring strategies during tutoring might have put a fairly high burden on a tutor's cognitive capacity (Feldon, 2007). Thus, there might not have been enough cognitive capacity left to derive information from a tutee's utterances of knowledge deficits as a basis for assessing a tutee's understanding after tutoring.

This interpretation is in accordance with results from research on the acquisition of memory strategies. Often, learners can spontaneously implement a newly learned memory strategy but experience a so-called *utilization deficiency* (Miller, 1990). That is, implementing the strategy does not immediately improve recall or even hinders it. It is argued that using a newly learned strategy, which is not yet automated, demands most of the cognitive capacity of a learner. Thus, there is little capacity left to spend on processing the material to be recalled (e.g., Miller & Seier, 1994).

Given this interpretation, it seems to be important to develop training methods that increase the automaticity with which interactive tutoring strategies are executed (Klauer, 1988). When interactive tutoring strategies occur more automatically, there might be more cognitive capacity available that can be used by tutors to assess a tutee's understanding (Feldon, 2007). Future research is encouraged to test whether training methods that target the automaticity of interactive tutoring strategies in fact improve assessment accuracy.

Acknowledgments

We thank Hannah Bartels, Victoria Denise Claes, Julian Etzel, Sophia Kammer, Rico Krieger, Amelie Krug, Annette Lehmann, Karina Meyer, Bosse Nietsch, Anne-Kristin Rückert, Tatjana Scharping, Eva Wiemers, Lisa Zimmer, and Raoul Zimmermann for their help with many practical aspects of the project. This research was supported by grants from the German Science Foundation DFG (WI 3348/2-1).

References

Brown, A. L., Campione, J. C., & Day, J. D. (1981). Learning to learn: On training students to learn from texts. *Educational Researcher, 10,* 14-21.

Cade, W. L., Copeland, J. L., Person, N. K., & D'Mello, S. K. (2008). *Dialogue modes in expert tutoring.* Paper presented at the Ninth International Conference on Intelligent Tutoring Systems, Montreal, Canada.

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

Chi, M. T. H., Roy, M., & Hausmann, R. G. M. (2008). Observing tutorial dialogues collaboratively: Insights about human tutoring effectiveness from vicarious learning. *Cognitive Science, 32*, 301-341.

Chi, M. T. H., Siler, S. A., & Jeong, H. (2004). Can tutors monitor students' understanding accurately?. *Cognition and Instruction, 22*, 363-387.

Chi, M. T. H., Siler, S. A., Jeong, H., Yamauchi, T., & Hausmann, R. G. M. (2001). Learning from human tutoring. *Cognitive Science, 25*, 471-533.

Cromley, J. G., & Azevedo, R. (2005). What do reading tutors do? A naturalistic study of more and less experienced tutors in reading. *Discourse Processes, 40*, 83-113.

Dignath, C., Buettner, G., & Langfeldt, H. P. (2008). How can primary school students learn self-regulated learning strategies most effectively? A meta-analysis on self-regulation training programmes. *Educational Research Review, 3*, 101-129.

D'Mello, S., Lehman, B. A., & Person, N. K. (2010). *Expert tutors feedback is immediate, direct, and discriminating.* Paper presented at the Twenty-Third International Florida Artificial Intelligence Research Society Conference (FLAIRS 2010), Key West, FL.

Feldon, D. F. (2007). Cognitive load and classroom teaching: The double-edged sword of automaticity. *Educational Psychologist, 42*, 123-137.

Graesser, A. C., D'Mello, S., & Cade, W. L. (2011). Instruction based on tutoring. In R. E. Mayer & P. A. Alexander (Eds.), *Handbook of research on learning and instruction*. New York: Routledge.

Hayes, A. F. (2009). Beyond Baron and Kenny: Statistical mediation analysis in the new millennium. *Communication Monographs, 76*, 408-420.

Hayes, A. F. (2012). *PROCESS: A versatile computational tool for observed variable mediation, moderation, and conditional process modeling* [White paper]. Retrieved from http://www.afhayes.com/public/process2012.pdf

Herppich, S., Wittwer, J., Nückles, M., & Renkl, A. (2013a). *Addressing knowledge deficits in tutoring and the role of teaching experience: Benefits for learning and summative assessment.* Manuscript submitted for publication.

Herppich, S., Wittwer, J., Nückles, M., & Renkl, A. (2013b). Does it make a difference? Investigating the assessment accuracy of teacher tutors and student tutors. *The Journal of Experimental Education, 81*, 242-260.

Hmelo-Silver, C. E., & Barrows, H. S. (2006). Goals and strategies of a problem-based learning facilitator. *Interdisciplinary Journal of Problem-based Learning, 1*, 21-39.

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

King, A., Staffieri, A., & Adelgais, A. (1998). Mutual peer tutoring: Effects of structuring tutorial interaction to scaffold peer learning. *Journal of Educational Psychology, 90*, 134-152.

Klauer, K. J. (1988). Teaching for learning-to-learn: A critical appraisal with some proposals. *Instructional Science, 17*, 351-367.

Leutner, D., Leopold, C., & Den Elzen-Rump, V. (2007). Self-regulated learning with a text-highlighting strategy: A training experiment. *Zeitschrift für Psychologie/ Journal of Psychology, 215*, 174-182.

Mandl, H., & Friedrich, H. F. (Eds.). (1992). *Lern- und Denkstrategien. Analyse und Intervention* [Learning and thinking strategies. Analysis and intervention]. Göttingen: Hogrefe.

McArthur, D., Stasz, C., & Zmuidzinas, M. (1990). Tutoring techniques in algebra. *Cognition and Instruction, 7*, 197-224.

Michael, J. A., Wenderoth, M. P., Modell, H. I., Cliff, W., Horwitz, B., McHale, P., … Whitescarver, S. (2002). Undergraduates' understanding of cardiovascular phenomena. *Advances in Physiology Education, 26*, 72-84.

Miller, P. H. (1990). The development of strategies of selective attention. In D. F. Bjorklund (Ed.), *Children's strategies: Contemporary views of cognitive development.* Hillsdale: Erlbaum.

Miller, P. H., & Seier, W. L. (1994). Strategy utilization deficiencies in children: When, where, and why. *Advances in Child Development and Behavior, 25*, 107-156.

Pelaez, N. J., Boyd, D. D., Rojas, J. B., & Hoover, M. A. (2005). Prevalence of blood circulation misconceptions among prospective elementary teachers. *Advances in Physiology Education, 29*, 172-181.

Perie, M., Marion, S., & Gong, B. (2009). Moving toward a comprehensive assessment system: A framework for considering interim assessments. *Educational Measurement: Issues and Practice, 28*, 5-13.

Preacher, K. J., & Kelley, K. (2011). Effect size measures for mediation models: Quantitative strategies for communicating indirect effects. *Psychological Methods, 16*, 93-115.

Renkl, A. (2005). The worked-out examples principle in multimedia learning. In R. E. Mayer (Ed.), *Cambridge handbook of multimedia learning.* New York: Cambridge University Press.

Roscoe, R. D., & Chi, M. T. H. (2007). Understanding tutor learning: Knowledge-building and knowledge-telling in peer tutors' explanations and questions. *Review of Educational Research, 77*, 534-574.

Shrout, P. E., & Bolger, N. (2002). Mediation in experimental and nonexperimental studies: New procedures and recommendations. *Psychological Methods, 7*, 422-445.

Sungur, S., & Tekkaya, C. (2003). Students' achievement in human circulatory system unit: The effect of reasoning ability and gender. *Journal of Science Education and Technology, 12*, 59-64.