Expertise Amiss: Interactivity Fosters Learning but Expert Tutors are Less Interactive than Novice Tutors

Stephanie Herppich, Jörg Wittwer, Matthias Nückles, and Alexander Renkl

University of Freiburg

Author Note

Stephanie Herppich, Jörg Wittwer, and Matthias Nückles, Department of Educational Science, University of Freiburg, Rempartstr. 11, D-79098 Freiburg, Germany; Alexander Renkl, Department of Psychology, University of Freiburg, Engelbergerstr. 41, D-79106 Freiburg, Germany.

Analyses in this article are based on data collected but not reported in Herppich et al. (2013, 2014). This research was supported by grants from the German Research Foundation (WI 3348/2-1).

Correspondence concerning this article should be addressed to Stephanie Herppich, e-mail: stephanie.herppich@ezw.uni-freiburg.de; telephone: +49 (0)761 203 2457; fax: +49 (0) 761 203 2458.

Abstract

The extent to which tutors are interactive and engage in dialogue with a student tends to depend on their pedagogical expertise. Normally, tutors with pedagogical expertise are more interactive than tutors without pedagogical expertise. This finding, however, has largely been obtained when examining tutoring in procedural domains such as mathematics. Hence, less is known about the extent to which tutors engage in interactivity as a function of their pedagogical expertise when they tutor a conceptual domain such as biology. Therefore, we conducted a study with $N = 46$ tutors who differed in their pedagogical expertise and examined their interactive style of tutoring in a conceptual domain. The results showed that a tutor's interactivity resulting from combining more scaffolding with less explaining particularly promoted a student's deep learning. Contrary to prior research, however, tutors with more pedagogical expertise were less interactive and, consequently, fostered learning to a lesser degree than tutors with less pedagogical expertise. Our findings suggest that a more complete understanding of interactivity in tutoring requires a differentiated approach considering interactivity as a multifaceted phenomenon.

Keywords: One-on-one human tutoring; instructional strategies; pedagogical expertise; learning

Introduction

In one-on-one human tutoring, one more knowledgeable tutor usually teaches one less knowledgeable student. Tutoring can, thus, be seen as the blueprint of instruction that draws on discourse between an instructor and a student. Accordingly, research has attributed the effectiveness of tutoring to those activities that are facilitated by the one-on-one discourse (Graesser et al. 1995). A student is deemed to learn best when a tutor engages in an interactive dialogue that elicits constructive responses from a student (Chi and Wylie 2014).

The ability to be interactive in tutoring seems to depend on a tutor's pedagogical expertise (Graesser et al. 2011). Research suggests that tutors with more pedagogical expertise are more interactive than tutors with less pedagogical expertise (Chae et al. 2005; Cromley and Azevedo 2005). However, most of the empirical work has exclusively examined tutoring in procedural domains such as mathematics. Therefore, it is an open question as to whether tutors with more pedagogical expertise are also more interactive when they tutor a conceptual domain such as biology.

To address this question, we examined the role of a tutor's interactivity for a student's learning of a conceptual domain as a function of a tutor's pedagogical expertise. Specifically, we investigated two types of interactivity that have been discussed as indices of a tutor's pedagogical expertise (Chi et al. 2008).

Interactivity in Tutoring

At the heart of one-on-one tutoring is the communication between a tutor and a student. Analyses of tutoring dialogues have shown that a tutor usually controls the communication (e.g., Chi et al. 2001, 2008; Graesser et al. 1995; Muldner et al. 2014). Therefore, the extent to which a tutor contributes to a student's learning depends on how a tutor guides the dialogue (Chi and Wylie 2014). In general, instruction is effective when learners get the opportunity to construct understanding that goes beyond the content of the information presented (VanLehn 2011). Therefore, tutoring should be most beneficial for learning when a tutor is interactive and guides the dialogue in a way that elicits meaningful responses from a student (Chi et al. 2008; Chi and Wylie 2014; Core et al. 2003; Jackson et al. 2004).

To be interactive in tutoring, a tutor might abstain from dominating the dialogue and exert only very gentle guidance. In this case, tutors would reduce their own contributions to the dialogue and, instead, overall, elicit more statements from a student (Chi et al. 2001). In doing so, students would be provided with the opportunity to make a larger number of meaningful contributions, thereby increasing the chance to improve their understanding (Chi et al. 2008).

Another way for a tutor to be interactive in tutoring would be to adopt specific instructional strategies. Two widely used instructional strategies in tutoring are explaining and scaffolding (e.g., Cromley and Azevedo 2005; Graesser et al. 1995). By *explaining* the tutor provides information that is relevant to the current content of the tutoring dialogue (Chi et al. 2008; D'Mello et al. 2010). However, explanations usually do not elicit responses from a student (Chi et al. 2001; Lu et al. 2007; VanLehn et al. 2003) and can discourage a student from being constructive when they are not adapted to a student's understanding (Wittwer and Renkl 2008). Therefore, explanations are not an interactive instructional strategy and seldom foster deep learning (Chi et al. 2001, 2008; Jackson et al. 2004). The term *scaffolding* is employed to describe a tutor's prompts, hints, and questions that are meant to help a student proceed "further in a line of reasoning or in a task that a tutee would not be able to accomplish alone." (Herppich et al. 2014, p. 936; cf. Chi et al. 2001). As scaffolding aims at eliciting constructive responses from a student, it is an interactive instructional strategy. In comparison to tutors who just reduce their own contributions in order to elicit statements from a student, a tutor who scaffolds, thus, purposefully aims at helping a student to achieve understanding of a specific content by being interactive. Consequently, research has found that scaffolding usually promotes learning (e.g., Chi and Wylie 2014; Core et al. 2003; Muldner et al. 2014).

In line with the two possibilities of being interactive in tutoring, Chi et al. (2008) proposed two measures of interactivity. The first measure concerns the number of a tutor and a student's contributions. When the number of a tutor's contributions is lower than the number of a student's contributions, a tutor would be less prone to dominate the dialogue and, thus, more interactive. The second measure refers to a tutor's explanations and scaffoldings. A tutor who explains less and scaffolds more would be more interactive than a tutor who explains more and scaffolds less.

Tutors' Pedagogical Expertise, Interactivity in Tutoring, and Students' Learning

Usually, tutors are profoundly knowledgeable about a subject matter but differ in terms of pedagogical training, pedagogical qualification, teaching experience, or tutoring experience (e.g., Cade 2009; Chi et al. 2001; 2008; Cromley and Azevedo 2005). To capture the level of a tutor's pedagogical expertise, Chi et al. (2008) suggested using the two measures of interactivity introduced in the previous section as indices of pedagogical expertise. For both measures, a higher level of interactivity would mirror a higher level of pedagogical expertise (see also Di Eugenio et al. 2006; Glass et al. 1999; Litman et al. 2006).

Research demonstrates that tutoring dialogues of pedagogical experts are in fact often more interactive than those of pedagogical novices (for an overview, see VanLehn 2011). For example, students tutored by pedagogical experts contribute to the dialogue to a greater extent than students tutored by pedagogical novices (e.g., Chi et al. 2001, 2008). In addition, more expert tutors predominantly scaffold their students whereas less expert tutors tend to explain a lot (e.g., Cromley and Azevedo 2005; Graesser et al. 1995).

To further corroborate that being highly interactive in tutoring is a sign of pedagogical expertise, pedagogical experts as tutors should support a student's learning more effectively than would pedagogical novices as tutors do. Evidence for this assumption, however, is scarce because many studies examining tutoring do not collect outcome measures (cf. Graesser et al. 2011). An exception is the study conducted by Chae et al. (2005) who investigated tutoring with a university lecturer as expert tutor and a university student as novice tutor. The results showed that the expert tutor elicited more contributions from the student and more often scaffolded the student than the novice tutor. Consequently, the expert tutor's students achieved larger learning gains than the novice tutor's students. Similarly, Herppich et al. (2014) contrasted teachers as tutors with pedagogical expertise against university students as tutors without pedagogical expertise. Their study revealed that teacher tutors more often engaged in scaffolding in response to a student's knowledge deficits than did student tutors. In doing so, the teacher tutors supported a student's conceptual learning to a greater extent than the student tutors.

Lu et al. (2007; see also Di Eugenio et al. 2006) examined an expert tutor with ample tutoring experience, a university lecturer as tutor with university classroom teaching experience, and a university student as novice tutor. Their results showed that students tutored by the expert tutor performed better on a posttest than students tutored by the two other groups of tutors. In contrast to the studies conducted by Chae et al. (2005) and Herppich et al. (2014), the expert tutor, however, predominantly used instructional strategies, like explaining, that do not actively elicit responses from a student. The lecturer tutor often scaffolded the student to be active, whereas the novice tutor mainly explained parts of the solution. Thus, the results of this study showed that even though the expert tutor supported a student's learning effectively he was not necessarily more interactive than the novice tutor.

Taken together, the reported studies (Chae et al. 2005; Chi et al. 2001, 2008; Cromley and Azevedo 2005; Graesser et al. 1995; Herppich et al. 2014) suggest that pedagogical experts as tutors are highly interactive in tutoring whereas pedagogical novices as tutors less often interact with their students. There are only few studies that do not confirm this positive relationship between interactivity and pedagogical expertise (Di Eugenio et al. 2006; Lu et al. 2007). In addition, research shows that pedagogical experts as tutors usually provide more effective tutoring than pedagogical novices as tutors. VanLehn (2011) presented some studies in his review, however, that failed to provide empirical evidence for an advantage of being tutored by a pedagogical expert. Yet, these studies did not examine a tutor's interactivity (e.g., Fossati, 2008; but see also the lecturer tutor in Lu et al. 2007, who had university classroom teaching experience and was interactive but did not promote learning more so than the novice tutor).

Tutors' Pedagogical Expertise, Interactivity in Tutoring, and Tutoring Domain

A closer inspection of the research reported in the previous section reveals that the majority of studies examined tutoring in a procedural domain where tutoring was mainly concerned with problem solving (e.g., Chae et al. 2005: algebra; Chi et al. 2008: physics; Cromley and Azevedo 2005: reading; Graesser et al. 1995: algebra, statistics). Hence, the positive relationship between interactivity and pedagogical expertise has primarily been documented when tutoring covered a procedural domain. Therefore, it is an open question as to whether tutors with more pedagogical expertise are also more interactive when they tutor a conceptual domain such as biology.

Due to the limited research on tutoring in conceptual domains with pedagogical experts as tutors, there is not much empirical evidence for possible differences in interactivity as a function of the type of domain covered in tutoring. An exception is the study conducted by Cade (2009) who examined the strategies of experienced professional tutors with a teaching license in mathematics (i.e., in a procedural domain) and in biology or in chemistry (i.e., in a conceptual domain). The study showed that expert tutors in mathematics mainly scaffolded their students whereas expert science tutors more intensively engaged in lecturing and, thus, in explaining. As previously reported, Herppich et al. (2014) found that tutors with pedagogical expertise more often engaged in scaffolding than tutors without pedagogical expertise. This result was obtained in a conceptual domain but referred exclusively to episodes in the tutoring sessions where tutors responded to a student's knowledge deficits (e.g., misconceptions). Whether pedagogical experts as tutors in general provided a more interactive style of tutoring than pedagogical novices as tutors was not reported in this study.

In summary, the two studies (Cade 2009; Herppich et al. 2014) provide mixed evidence as to whether tutors with pedagogical expertise are more interactive than tutors without pedagogical expertise when tutoring a conceptual domain. Due to the nature of a conceptual domain, one might, however, expect that tutors in fact engage more intensively in explaining than in eliciting problem-solving activities from a student (for a similar discussion, see VanLehn 2011). Still, interactivity should play an important role for the effectiveness of tutoring in conceptual domains, too. For example, Chi et al. (2001) showed that learning from tutoring in a conceptual domain was influenced by the extent to which tutors were interactive.

Present Study and Hypotheses

In this article, we present a study in which we examined the extent to which tutors with varying levels of pedagogical expertise were interactive when tutoring a conceptual domain and, thereby, supported a student's learning. In our study, we aimed to overcome some of the limitations of previous research on tutoring. First, only few studies investigated a tutor's activities and a student's learning as a function of a tutor's level of pedagogical expertise (Chae et al. 2005; Herppich et al. 2014; Lu et al. 2007; cf. Graesser et al. 2011; VanLehn 2011). Thus, the aim of our study was to systematically compare tutors with varying pedagogical expertise with regard to their interactivity and their effectiveness in terms of a student's learning. Second, the majority of studies analyzed tutoring in procedural domains (e.g., Chae et al. 2005). Hence, our study examined whether the findings obtained in these studies would be generalizable to a conceptual domain. Third, the different approaches to being interactive in tutoring as suggested by the two measures of interactivity proposed by Chi et al. (2008) have not been systematically tested yet. Therefore, we examined these measures in more detail to find out whether they are in fact valid indices of a tutor's pedagogical expertise.

The protocol data used for the analyses presented in this article were collected in the study reported by Herppich et al. (2013, 2014) but were not examined with regard to a tutor's overall interactivity. More specifically, in Herppich et al. (2013), we investigated a tutor's competence at assessing a student's learning after tutoring. In Herppich et al. (2014), we specifically focused on the knowledge deficits (e.g., misconceptions) that students uttered in the course of tutoring and how tutors responded to these knowledge deficits. In response to a knowledge deficit, tutors, for example, might have been interactive and scaffolded a student to learn more about the knowledge deficit. In the present study, however, we were interested in the interactivity that tutors engaged in across the complete course of tutoring. Hence, in contrast to the fine-grained level of knowledge deficits examined in Herppich et al. (2014), the present study investigated interactivity at the coarse-grained level of a whole tutoring session.

We derived the following hypotheses from the premise that interactivity would foster learning and from the premise that tutors with more pedagogical expertise would be more interactive than tutors with less pedagogical expertise:

1. Interactivity-Learning Hypothesis: The more a tutor engages in interactivity the more a student learns from tutoring.
2. Expertise-Interactivity Hypothesis: Tutors with more pedagogical expertise are more interactive in tutoring than tutors with less pedagogical expertise.
3. Mediation Hypothesis: Tutors with more pedagogical expertise promote a student's learning more effectively than tutors with less pedagogical expertise because expert tutors are more interactive than novice tutors.

Method

A more detailed description of the method can be found in Herppich et al. (2013).

Sample and Design

Overall, *N* = 46 tutor–student dyads participated in the study. Twenty-one biology teachers with a mean age of 44.05 years (*SD* = 11.76) served as tutors with pedagogical expertise (= *teacher tutors*) in this study. Eleven teacher tutors were female. Twenty-five university students of biology without professional teaching experience with a mean age of 22.24 years (*SD* = 2.83) served as tutors without pedagogical expertise (= *student tutors*) in this study. Twenty-one student tutors were female. Teacher tutors and student tutors had comparable tutoring experience, $U = 248.00$, $z = -0.34$, $p = .73$, $f = 0.05$ (small effect; tutoring experience coded as 1 = *no experience*, 2 = *occasional tutoring*, 3 = *regular tutoring*) and comparable subject matter knowledge (see Herppich et al. 2013). The students being tutored in this study attended the seventh grade. Their mean age was 12.65 years (*SD* = 0.53). Nineteen students were female. They were randomly assigned to one of the two tutor groups. As detailed in Herppich et al. (2013), all students had a comparably low level of knowledge about the human circulatory system. The dependent variables were (1) a tutor's interactivity during the tutoring dialogue and 2) a student's learning gain from the beginning to the end of the tutoring dialogue.

Materials

Textbook passage. During tutoring, the tutor–student dyads worked through a passage about the human circulatory system (Towle 1989; cf. Chi et al. 2001, 2004; Herppich et al. 2013).

Concepts test. The test consisted of 25 multiple-choice items that assessed a student's understanding of single concepts related to the human circulatory system. The reliability of the test was acceptable (see Herppich et al. 2013). A correct answer indicated a correct understanding. Each of the incorrect answers indicated an incorrect understanding that had been identified as a misconception by previous research (e.g., Pelaez et al. 2005; for more information, see Herppich et al. 2013). Each item that a student answered correctly was assigned 1 point. Hence, a student could achieve a maximum number of 25 points.

Mental model test. On a sheet of paper, the outline of a human body was displayed. The students drew the blood path of the circulatory system into the outline and explained the blood path. Explanations were audiotaped. By using this methodology, originally developed by Chi et al. (2004), we assessed a student's understanding about the human circulatory system at the level of mental models. To code the students' drawings and explanations, we adapted a classification scheme originally developed by Azevedo et al. (2004). Based on this classification scheme, the drawings and explanations were assigned a score between 0 and 11. The scores reflect different levels of understanding about the human circulatory system ranging from 0 (= *no understanding*) to 11 (= *complete understanding*; Azevedo et al. 2004). The intraclass correlations measuring absolute agreement between two coders were satisfying (see Herppich et al. 2013).

Procedure

Each tutoring session comprised three phases: pretest, tutoring, and posttest. On average, a tutoring session lasted about three hours. At pretest, the students completed the concepts test and the mental model test. Additionally, each student and each tutor read the passage about the human circulatory system. In the tutoring phase, tutor–student dyads jointly read the passage about the human circulatory system sentence-by-sentence and engaged in a dialogue about each sentence. The tutoring phase was videotaped. At posttest, the students again completed the concepts test and the mental model test.

Coding

Statements and instructional strategies. To code the number of statements uttered by a student and uttered by a tutor and to code a tutor's instructional strategies from the transcribed tutoring dialogues, we adapted the statement coding system developed by Chi et al. (2001). First, a coder segmented those parts of the transcripts into statements in which the tutor and the student discussed the human circulatory system. We did not segment or code speech irrelevant to the subject matter such as small talk. A statement contains a single idea (cf. Chi et al. 2001; for examples, see Table 1). The segmentation was ‘based on the structure of speech’ (Chi et al. 2008, p. 322). Where the segmentation based on structural markers was too coarse or did not produce meaningful statements, we refined it on the basis of grammatical and organizational markers (e.g., *and*, *or*, *because*, *for example*, *such as*, *that is*; Wittwer et al. 2010, p. 80). On average, a tutor uttered 550 (SD = 245) statements and a student uttered 260 (SD = 132) statements.

Second, to classify the resulting substantive statements, the coder assigned each statement of a tutor to one of eight categories. These are (1) giving explanations, (2) giving direct positive or negative feedback, (3) reading text sentences aloud, (4) making self-monitoring comments, (5) answering questions that the student asked, (6) asking content questions, (7) scaffolding, (8) asking comprehension-gauging questions. As displayed in Table 2, tutors used explaining and scaffolding more often than all other instructional strategies. This difference was statistically significant, $F(1, 45) = 139.67$, $p < .001$, $f = 1.76$ (large effect). In addition, none of the instructional strategies significantly correlated in isolation with a student’s learning.

To standardize segmentation and coding, the coder used a detailed code book. A second coder double coded 10 transcripts, comprising 8622 statements. The interrater agreement was $\kappa = .96$ (Cohen 1960) for identifying a statement as being uttered by a tutor, $\kappa = .97$ for identifying a statement as being uttered by a student, $\kappa = .71$ for explaining, and $\kappa = .60$ for scaffolding. Thus, the interrater agreement for all codings was good (Fleiss and Cohen 1973).

Interactivity. We computed the two measures of a tutor’s interactivity as suggested by Chi et al. (2008). Due to being ratios, both measures correct for possible differences in dialogue-length between dyads. The *contribution-based interactivity* related the number of a student's contributions to the number of a tutor's contributions. We summed up all statements uttered, separately for each tutor and for each student, and divided the sum of a student's statements by the sum of a tutor's statements for each tutor-student dyad. A value above 1 indicated a higher level of contribution-based interactivity because the student uttered more statements than the tutor. Hence, in this case, the tutor more often allowed for a student's participation. A value below 1 indicated a lower level of contribution-based interactivity because the tutor uttered more statements than the student.

Second, the *instruction-based interactivity* was derived from the ratio of a tutor's scaffoldings to a tutor's explanations. We summed up the number of a tutor's statements coded as scaffolding and the number of a tutor's statements coded as explaining separately for each tutor. Then, we divided the sum for scaffolding by the sum for explaining for each tutor. A value above 1 indicated a higher level of instruction-based interactivity because scaffolding occurred more often than explaining, whereas a value below 1 indicated a lower level of instruction-based interactivity because explanations were more prevalent than scaffolding.

Students' learning gain. We used two measures to assess a student's learning gain. First, we measured the learning of concepts by subtracting the pretest score in the concepts test from the posttest score in the concepts test. To avoid bias in favor of students with lower knowledge, we adjusted a student's learning gain for the possible maximum learning gain. Thus, a student who had 4 points in the pretest and 18 points in the posttest achieved a learning gain for concepts of (18 – 4)/(25 – 4) = .67 (Herppich et al. 2014). Second, we proceeded in the same way to measure a student's learning gain for mental models. For example, if a student's drawings and explanations were assigned a score of 3 in the pretest and a score of 9 in the posttest, the learning gain for mental models was: (9-3)/(11-3) = .75.

Results

For all directional hypotheses, we used one-tailed tests. In addition to *p*-values, we report 95% bias-corrected and accelerated bootstrapping confidence intervals (CI_B) based on

1,000 resamples with replacement to test the hypotheses (DiCiccio and Efron 1996). We used f (Cohen 1988) as effect-size measure for all analyses. Cohen (1988) suggested that f = 0.10 is a small effect, f = 0.25 is a medium effect, and f = 0.40 is a large effect. Only for indirect effects in simple mediation analyses, we employed κ^2 as effect size measure. Preacher and Kelley (2011) suggested to interpret $\kappa^2 = .01$ as small effect, $\kappa^2 = .09$ as medium effect, and $\kappa^2 = .25$ as large effect. All analyses were performed using Excel 2010, SPSS 22.0.0, and the PROCESS macro for SPSS (Hayes 2012).

Interactivity-Learning Hypothesis

The interactivity-learning hypothesis predicted that a tutor would foster a student's learning the more, the more the tutor was interactive. To test this hypothesis, we performed four simple linear regression analyses. We used each of the two measures of a tutor's interactivity, that is, contribution-based interactivity and instruction-based interactivity, as the predictor and a student's concept learning and a student's mental model learning, respectively, as the criterion (see Table 3 for means and standard deviations). The first regression analysis with contribution-based interactivity as the predictor and concept learning as the criterion was not significant, $R^2 < .01$, $F(1, 44) < 0.01$, $b < .01$, $p = .98$, CI_B [-.22, .32], $f < 0.01$ (small effect). The second regression analysis with the same predictor but with mental model learning as the criterion revealed a trend showing that this type of interactivity positively influenced a student's learning, $R^2 = .06$, $F(1, 44) = 2.55$, $b = .25$, p =.06, CI_B [.04, .54], $f = .24$ (small effect). The third regression analysis with instruction-based interactivity as the predictor and concept learning as the criterion was again not significant, $R^2 = .01$, $F(1, 44) = 0.29$, $b = .03$, $p = .59$, CI_B [-.06, .09], $f = 0.08$ (small effect). Finally, the fourth regression analysis, which tested the effect of instruction-based interactivity on mental model learning, indicated a significant effect, $R^2 = .07$, $F(1, 44) = 3.40$, $b = .13$, $p = .04$, CI_B [.04, .23], $f = 0.28$ (medium effect). In summary, the results partially supported our interactivity-learning hypothesis. Whereas interactivity did not substantially influence concept learning, it fostered mental model learning. This result was more pronounced for instruction-based interactivity than for contribution-based interactivity.

Expertise-Interactivity Hypothesis and Mediation Hypothesis

The expertise-interactivity hypothesis stated that teacher tutors would be more interactive in tutoring than student tutors. Moreover, the mediation hypothesis assumed that teacher tutors would support a student's learning more effectively than student tutors because teacher tutors should be more interactive than student tutors. To test both hypotheses, we performed four mediation analyses. We calculated total, direct, and indirect effects by applying regression-based path analyses. In each mediation analysis, the predictor was a tutor's level of pedagogical expertise. Teacher tutors were coded as 1 and student tutors were coded as 0. The mediator was a tutor's interactivity and the criterion was a student's learning. To test the indirect effects, we used the bootstrapping procedure as suggested by Preacher and Hayes (2008; see also Hayes 2012). The procedure was based upon 10,000 resamples with replacement to derive a 95% bias-corrected confidence interval for the indirect effects. We tested both hypotheses based on the same four analyses because the path from the predictor to the mediator equals a simple linear regression analysis. We used the same combination of variables as before in our statistical tests of the interactivity-learning hypothesis.

In the first and the second mediation analysis, contribution-based interactivity was the mediator (Fig. 1 a and b; see Table 3 for means, standard deviations, and correlations). Contrary to the prediction made in the expertise-interactivity hypothesis, we found a negative effect of pedagogical expertise on the degree of contribution-based interactivity. This effect, however, was not significant, $R^2 = .03$, $F(1, 44) = 1.55$, $p = .11$, CI_B [-0.25, 0.04], $f = 0.19$ (small effect). In the third and the fourth mediation analysis, the mediator was instruction-based interactivity (Fig. 1 c and d). In these analyses, again, the effect of pedagogical expertise on a tutor's interactivity was negative but, this time, it was significant, $R^2 = .16$, $F(1, 44) = 8.24$, $p < .01$, CI_B [-0.85, -0.22], $f = 0.43$ (large effect). Hence, instead of being more interactive, teacher tutors were less interactive because they engaged even more in explaining than in scaffolding than did student tutors.

Regarding the mediation hypothesis, the first mediation analysis tested the indirect effect of pedagogical expertise on a student's concept learning via contribution-based interactivity (Fig. 1 a). This mediation analysis revealed a negative unstandardized indirect effect of -.01 (κ^2 = .01, small effect), CI_B [-.05, .01]. This confidence interval includes zero. Thus, the indirect effect is inconsistent with the mediation hypothesis but small and not significant[1]. The second mediation analysis tested the indirect effect of pedagogical expertise on a student's mental model learning via contribution-based interactivity (Fig. 1 b). It yielded a significant negative unstandardized indirect effect of -.03 (κ^2 = .05, small effect), CI_B [.08, .0002]. In the third mediation analysis, we examined the effect of pedagogical expertise on a student's concept learning via instruction-based interactivity (Fig. 1 c). The indirect effect of -.05 (κ^2 = .10, medium effect), CI_B [-.10, -.003] was significant. Finally, the fourth mediation analysis investigated the effect of pedagogical expertise on a student's mental model learning (Fig. 1 d). Again, we found a negative unstandardized indirect effect. Its value was -.08 (κ^2 = .12, medium effect) and it was significant, CI_B [-.14, .03].

In summary, all four mediation analyses yielded indirect effects that were in contrast with our mediation hypothesis: Contrary to our prediction, the student tutors promoted a student's learning more strongly than the teacher tutors because they were more interactive than the teacher tutors. This pattern of result was obtained in three out of four mediation analyses.

[1] Fig. 1 (a and c) also displays the total effect of a tutor's expertise on a student's concept learning that was reported in Herppich et al. (2014). Coefficients deviate marginally from those reported earlier. This is because in Herppich et al (2014), one tutor-student-dyad did not yield codes for the analyses performed there and, thus, had to be excluded from the analysis. For the analyses presented in this article, all dyads yielded codes and, accordingly, could be included. The total effect of the type of tutor on a student' concept learning was significant, $R^2 = .09$, $F(1,44) = 4.26$, $p = .04$, 95% CI_B [.01; .28], $f = 0.31$ (medium effect). The total effect of the type of tutor on a student' mental model learning (Fig. 1 b and d), yet, was not significant, $R^2 < .01$, $F(1,44) < 0.01$, $p = .96$, CI_B [-.16, .18], $f = 0.01$ (small effect).

Discussion

In this study, we examined the role of a tutor's interactive style of tutoring for a student's learning as a function of a tutor's level of pedagogical expertise. First, we found that being more interactive as tutor generally helped a student to learn the mental model of the human circulatory system, which is in line with previous research (e.g., Chi et al. 2001, 2008; Chi and Wylie 2014; Mulder et al. 2014). The students in our study were relatively young and had low prior knowledge (see also Herppich et al. 2013). Hence, our finding together with results from prior research (Muldner et al. 2014) suggests that tutor-guided interactive instruction is particularly beneficial for younger students and for students with low prior knowledge. This makes intuitively sense because these groups of students usually do not benefit optimally from less-structured learning environments (cf. VanLehn et al. 2007).

Second, interactivity failed to have a substantial impact on a student's learning of concepts. This might have been the case because our analysis of interactivity was on a rather coarse-grained level. That is, the analysis referred to activities that a tutor engaged in across the duration of tutoring. Accordingly, these activities should primarily help students to learn the structure and the function of the human circulatory system as a whole. The mental model test exactly measured such a general understanding of the human circulatory system. The concepts tests, in contrast, captured rather specific (mis-)understandings of concepts such as "What is the task of the heart in the human organism?" (Herppich et al. 2013, p. 249). Obviously, the acquisition of such specific pieces of knowledge was not directly influenced by a tutor's overall activities in tutoring. This assumption is corroborated by findings of Herppich et al. (2014). In this study, the authors performed a rather fine-grained analysis of a student's knowledge deficits and found that tutors supported a student's learning of concepts when they formatively assessed a student's knowledge deficits during tutoring.

Third, we found that the effect of interactivity on mental model learning was more pronounced for instruction-based interactivity as measured by the type of a tutor's instructional strategies than for contribution-based interactivity as measured by the relative amount of a tutor and a student's contributions. This result might be interpreted in the light of research on teaching effectiveness. In this research, there is a distinction between proximal and distal components of teaching (e.g., Seidel and Shavelson 2007). Teaching components that are more proximal to a student's execution of learning processes, such as providing learning opportunities to execute domain-specific learning activities, usually have larger positive effects on student outcomes than teaching components that are more distal to the execution of learning processes, such as setting the social context for learning. Analogously, a tutor's use of instructional strategies that support interactivity can be regarded as more proximal to a student's execution of learning processes than the mere amount of a student's contributions (cf. Litman et al. 2006). Thus, a type of interactivity that refers to a tutor's instructional strategies is obviously a stronger predictor for learning than a type of interactivity that primarily relies on the number of a tutor and a student's contributions to tutoring. Hence, with regard to the two measures of pedagogical expertise developed by Chi et al. (2008), our findings suggest that instruction-based interactivity is a more important (and, maybe, more valid) index of pedagogical expertise than contribution-based interactivity.

Fourth, and in contrast to prior research, this study revealed that tutors with pedagogical expertise were less interactive than tutors without pedagogical expertise. As a consequence of being less interactive, pedagogical experts as tutors promoted a student's learning to a lesser extent than did pedagogical novices as tutors. This unexpected finding might be explained by the fact that we examined tutoring in a conceptual domain. Like the science tutors in the study conducted by Cade (2009), our biology teacher tutors tended to use non-interactive strategies such as explaining more often than interactive strategies such as scaffolding (however, note that student tutors were not optimally interactive, either, cf. Table 3). Possibly, this is because it may be more natural in conceptual domains to give explanations instead of eliciting responses from students (VanLehn 2011). Why the teacher tutors used interactive strategies less often than the student tutors is open to interpretation. One explanation is that experienced teachers, more so than university students, are especially prone to 'lecture' in a conceptual domain. Although it might be a suboptimal way to help students to acquire understanding, it may be the form of teaching they are used to (cf. Hattie 2012). To get a better understanding about the role of the domain for a tutor's interactivity, future research is encouraged to directly compare tutoring in different domains (i.e., procedural and conceptual domain) with tutors who differ in their pedagogical expertise.

Fifth, that pedagogical experts as tutors were overall less interactive than novice tutors stands in contrast to the findings obtained by Herppich et al. (2014). In their analysis, the same experts were more interactive with regard to formatively assessing a student's knowledge deficits than the same novices. This differing pattern of results suggests that pedagogical experts as tutors in a conceptual domain in general have a less interactive style of tutoring than pedagogical novices as tutors. However, in specific situations of a tutoring session, such as when a student expresses a knowledge deficit, pedagogical experts as tutors seem to engage more intensively in interactivity than pedagogical novices as tutors to learn more about a student's individual understanding, which in turn supports learning (see Fig. 1 a c; Herppich et al. 2014).

Combined, the present study and the Herppich et al. (2014) study, show that the extent to which the same tutors engage in interactivity and the extent to which this interactivity supports learning differ depending on the level of analysis. This finding makes a strong case for requiring a more differentiated approach to analyzing tutoring dialogues. Obviously, tutoring dialogues seem to be a complex amalgam. This amalgam renders simple conclusions about the influence of interactivity on learning or about differences in interactivity between pedagogical experts as tutors and pedagogical novices as tutors difficult. Thus, research needs to ask 'which interactivity', 'which learning', and 'which content' to draw more differentiated conclusions about what characterizes competent or expert tutoring.

A limitation of the present study might be that our analyses focused on a tutor’s activities. It can be argued that a student’s contributions are more important for learning than a tutor’s contributions (Chi et al. 2001, 2008; Muldner et al. 2014) because a student’s actions are more intimately linked to learning. Nevertheless, students often contribute to the tutoring not until they are prompted by the tutor to do so (e.g., Muldner et al. 2014). This is particularly true for younger students and for students with low prior knowledge such as those students in our study (cf. Muldner et al. 2014, VanLehn et al. 2007). As a student’s contributions in this case cannot be considered isolated from the quality of a tutor’s contributions, it is certainly important what tutors do to support a student’s learning.

References

Azevedo, R., Cromley, J. G., & Seibert, D. (2004). Does adaptive scaffolding facilitate students' ability to regulate their learning with hypermedia?. *Contemporary Educational Psychology, 29*(3), 344-370. doi:10.1016/j.cedpsych.2003.09.002.

Cade, W. L. (2009). *How domain differences impact the mode structure of expert tutoring dialogue*. Honors paper. Rhodes College, Memphis, TN. www.researchgate.net.

Chae, H. M., Kim, J. H., & Glass, M. (2005). Effective behaviors in a comparison between novice and expert algebra tutors. In S. Hettiarachchi & R. Finkbine (Eds.), *Proceedings of the Sixteenth Midwest Artificial Intelligence and Cognitive Science Conference*. http://www.mglass.org/papers/papers.html.

Chi, M. T. H., & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist, 49*(4), 219-243. doi: 10.1080/00461520.2014.965823.

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*(2), 301-341. doi: 10.1080/03640210701863396.

Chi, M. T. H., Siler, S. A., Jeong, H., Yamauchi, T., & Hausmann, R. G. M. (2001). Learning from human tutoring. *Cognitive Science: A Multidisciplinary Journal, 25*(4), 471-533. doi:10.1207/s15516709cog2504_1.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20*(1), 37-46. doi:10.1177/001316446002000104.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciencies* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Core, M. G., Moore, J. D., & Zinn, C. (2003). The role of initiative in tutorial dialogue. In A. Copestake & J. Hajic (Eds.), *Proceedings of the 10th Conference of the European*

Chapter of the Association for Computational Linguistics (pp. 67-74). www.aclweb.org.

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*(2), 83-113. doi: 10.1207/s15326950dp4002_1.

D'Mello, S., Lehman, B. A., & Person, N. K. (2010). Expert tutors feedback is immediate, direct, and discriminating. In H. W. Guesgen & C. Murray (Eds.), *Proceedings of the Twenty-Third International Florida Artificial Intelligence Research Society Conference* (pp. 504-509). Menlo Park, CA: AAAI Press. http://www.aaai.org.

DiCiccio, T. J., & Efron, B. (1996). Bootstrap confidence intervals. *Statistical Science, 11*(3), 189-212. doi:10.1214/ss/1032280214.

Di Eugenio, B., Kershaw, T. C., Lu, X., Corrigan-Halpern, A., & Ohlsson, S. (2006). Toward a computational model of expert tutoring: a first report. In G. Sutcliffe, & R. Goebel (Eds.), *Proceedings of 19th International conference of Florida Artificial Intelligence Research Society* (pp. 503-508). Menlo Park, CA: AAAI Press. http://www.aaai.org.

Fleiss, J. L., & Cohen, J. (1973). The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educational and Psychological Measurement, 33*(3), 613-619. doi:10.1177/001316447303300309.

Glass, M., Kim, J. H., Evens, M. W., Michael, J. A., & Rovick, A. A. (1999). Novice vs. expert tutors: A comparison of style. In U. Priss (Ed.), *Midwest Artificial Intelligence and Cognitive Science Conference* (pp. 43-49). Bloomington, IN: AAAI Press.

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* (2nd ed., pp. 408-426). London, United Kingdom: Routledge.

Graesser, A. C., Person, N. K., & Magliano, J. P. (1995). Collaborative dialogue patterns in naturalistic one-to-one tutoring. *Applied Cognitive Psychology, 9*(6), 495-522. doi: 10.1002/acp.2350090604.

Hattie, J. (2012). *Visible learning for teachers. Maximizing impact on learning*. London: Routledge.

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. (2013). Does it make a difference? Investigating the assessment accuracy of teacher tutors and student tutors. *The Journal of Experimental Education, 81*(2), 242-260. doi:10.1080/00220973.2012.699900.

Herppich, S., Wittwer, J., Nückles, M., & Renkl, A. (2014). Addressing knowledge deficits in tutoring and the role of teaching experience: Benefits for learning and summative assessment. *Journal of Educational Psychology*, *106*(4), 934-945. doi:10.1037/a0036076.

Jackson, G., Person, N., Graesser, A. (2004). Adaptive tutorial dialogue in AutoTutor. In Lester, J. C., Vicari, R. M., Paraguaçu, F. (Eds.), *Proceedings of the workshop on dialog-based intelligent tutoring systems at the 7th International Conference on Intelligent Tutoring Systems* (Vol. 3220, pp. 368-372) Heidelberg, Germany: Springer. http:// ibrarian.net.

Litman, D. J., Rosé, C. P., Forbes-Riley, K., VanLehn, K., Bhembe, D., & Silliman, S. (2006). Spoken versus typed human and computer dialogue tutoring. *International Journal of Artificial Intelligence in Education, 16*(2), 145-170. http://ijaied.org/.

Lu, X., Di Eugenio, B., Kershaw, T., Ohlsson, S., & Corrigan-Halpern, A. (2007). Expert vs. non-expert tutoring: Dialogue moves, interaction patterns and multi-utterance turns. In

A. Gelbukh (Ed.), *Computational linguistics and intelligent text processing* (Vol. 4394, pp. 456-467). Heidelberg, Germany: Springer.

Muldner, K., Lam, R., & Chi, M. T. H. (2014). Comparing learning from observing and from human tutoring. *Journal of Educational Psychology, 106*(1), 69-85. doi: 10.1037/a0034448.

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*(3), 172–181. doi: 10.1152/advan.00022.2004.

Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods, 40*(3), 879-891. doi:10.3758/brm.40.3.879.

Preacher, K. J., & Kelley, K. (2011). Effect size measures for mediation models: Quantitative strategies for communicating indirect effects. *Psychological Methods, 16*(2), 93-115. doi:10.1037/a002265810.1037/a0022658.supp.

Seidel, T., & Shavelson, R. J. (2007). Teaching effectiveness research in the past decade: The role of theory and research design in disentangling meta-analysis results. *Review of Educational Research, 77*(4), 454-499. doi: 10.3102/0034654307310317.

Towle, A. (1989). *Modern biology*. New York, NY: Holt, Rinehart and Winston.

VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist, 46*(4), 197-221. doi:10.1080/00461520.2011.611369.

VanLehn, K., Graesser, A. C., Jackson, G. T., Jordan, P., Olney, A., & Rosé, C. P. (2007). When are tutorial dialogues more effective than reading?. *Cognitive Science: A Multidisciplinary Journal, 31*(1), 1-60. doi: 10.1080/03640210709336984.

VanLehn, K., Siler, S. A., Murray, C., Yamauchi, T., & Baggett, W. B. (2003). Why do only some events cause learning during human tutoring?. *Cognition and Instruction, 21*(3), 209-249. doi:10.1207/s1532690xci2103_01.

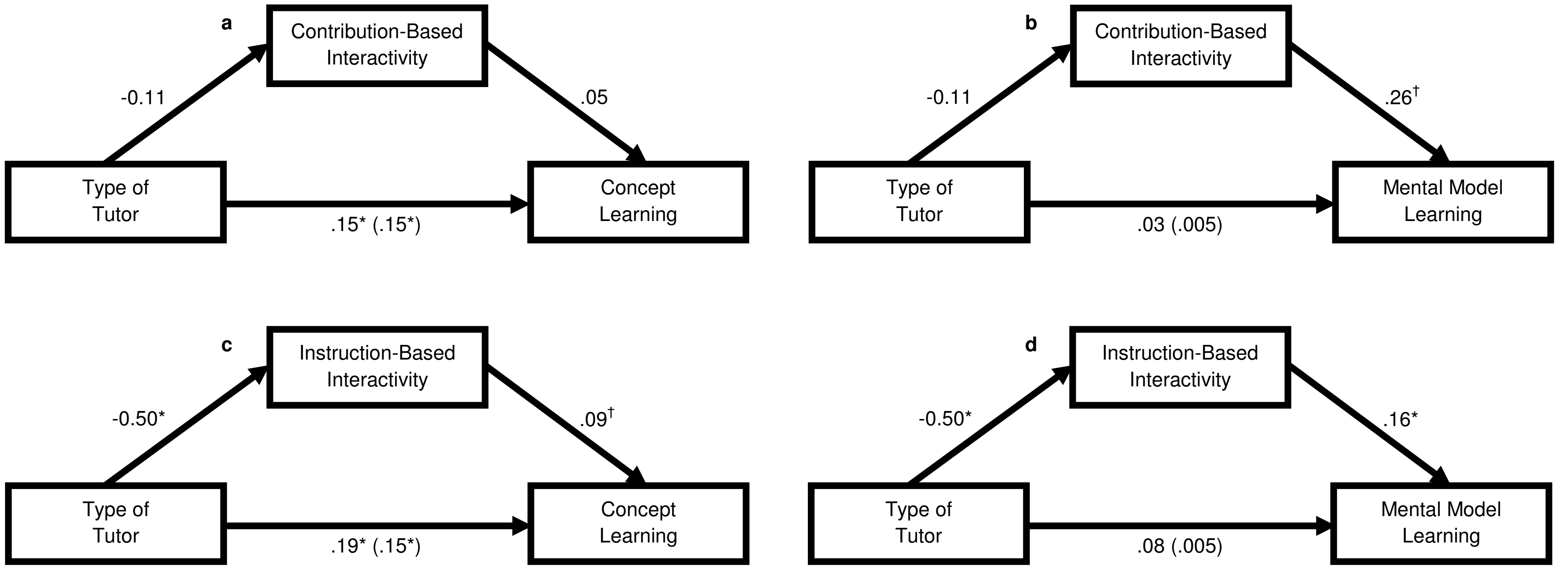
Wittwer, J., Nückles, M., Landmann, N., & Renkl, A. (2010). Can tutors be supported in giving effective explanations?. *Journal of Educational Psychology, 102*(1), 74-89. doi: 10.1037/a0016727.

Wittwer, J., & Renkl, A. (2008). Why instructional explanations often do not work: A framework for understanding the effectiveness of instructional explanations. *Educational Psychologist, 43*(1), 49-64. doi:10.1080/00461520701756420.

Figure Captions

Fig. 1

A tutor's contribution-based interactivity did not mediate the relationship of the type of tutor (i.e., teacher tutors and student tutors) with a student's concept learning (a). There was a trend for a mediation of the relationship of the type of tutor with a student's mental model learning (b). A tutor's instruction-based interactivity mediated the relationship of the type of tutor with a student's concept learning (c) and a student's mental model learning (d). Numbers represent unstandardized path coefficients for direct and, in parentheses, total effects. $^{*}p < .05$. $^{\dagger}p < .1$



a
Contribution-Based Interactivity
Type of Tutor
Concept Learning
-0.11
.05
.15* (.15*)
b
Contribution-Based Interactivity
Type of Tutor
Mental Model Learning
-0.11
.26†
.03 (.005)
c
Instruction-Based Interactivity
Type of Tutor
Concept Learning
-0.50*
.09†
.19* (.15*)
d
Instruction-Based Interactivity
Type of Tutor
Mental Model Learning
-0.50*
.16*
.08 (.005)

Table 1

Examples for the segmentation of transcripts (indicated by the % sign) and coding of a tutor's instructional strategies

Speaker	Utterances	Type of a tutor's strategy
Episode 1		
Student	[Reads a sentence] "The body's largest artery is the aorta. % Starting from the aorta, blood is transported through smaller arteries."%	
Tutor	Right – it's really thick – the aorta, and particularly stable.%	Explaining
Episode 2		
Student	[Reads a sentence] "The right side of the heart pumps the blood towards the lungs.% The left side of the heart pumps the blood towards all other parts of the body.%"	
Tutor	Okay, well, so do we only have one circuit in our body?%	Scaffolding

Table 2

Means and Standard Deviations of a Tutor's Instructional Strategies, Differences between Tutor Groups and Correlations with a Student's Learning

Strategy	Teacher tutors $n = 21$ M (SD)	Student tutors $n = 25$ M (SD)	Difference between tutor groups $F(1, 44)$ (f)	All tutors $N = 46$ M (SD)	Concept learning r	Mental model learning r
Explaining	.51 (.14)	.40 (.14)	7.27* (0.41)	.45 (.15)	.03	-.23
Feedback	.15 (.06)	.15 (.11)	0.06 (0.03)	.15 (.09)	.15	.20
Reading	.02 (.03)	.08 (.11)	5.87* (0.37)	.06 (.09)	-.22	-.05
Self-Monitoring	.03 (.02)	.03 (.03)	0.01 (0.01)	.03 (.02)	-.19	.03
Answering	.03 (.04)	.01 (.02)	2.88† (0.26)	.02 (.03)	.24	.05
Asking	.07 (.05)	.01 (.01)	29.98* (0.83)	.04 (.04)	.07	-.14
Scaffolding	.16 (.10)	.27 (.10)	14.55* (0.58)	.22 (.12)	-.02	.21
CGQs	.03 (.02)	.03 (.02)	0.64 (0.12)	.03 (.02)	-.16	-.04

Note. CGQs = comprehension-gauging questions, $^{*}p < .05$, $^{\dagger}p < .10$.

Table 3

Means, Standard Deviations, and Intercorrelations of the Number of Statements Uttered, a Tutors' Instructional Strategies and a Student's Learning Gains

Measure	Teacher tutors $n = 21$ M (SD)	Student tutors $n = 25$ M (SD)	All tutors $N = 46$ M (SD)	1	2	3	4	5
1. Type of tutor	—	—	—	—				
2. Contribution-based interactivity	.45 (.25)	.56 (.32)	.51 (.29)	-.18	—			
3. Instruction-based interactivity	.38 (.35)	.88 (.73)	.65 (.64)	-.40*	-.21	—		
4. Concept learning	.28 (.25)	.13 (.24)	.20 (.25)	-.30*	.01	.08	—	
5. Mental model learning	.60 (.35)	.60 (.28)	.60 (.31)	.01	.23	.27†	.21	—

Note. *$p < .05$, $^{\dagger}p < .10$.