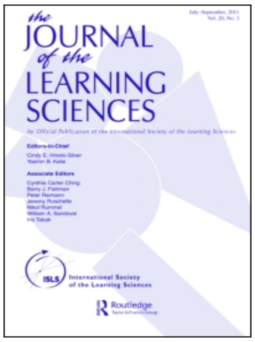

Routledge
Taylor & Francis Group

Journal of the Learning Sciences

ISSN: 1050-8406 (Print) 1532-7809 (Online) Journal homepage: https://www.tandfonline.com/loi/hlns20

Why Students Learn More From Dialogue- Than Monologue-Videos: Analyses of Peer Interactions

Michelene T. H. Chi, Seokmin Kang & David L. Yaghmourian

To cite this article: Michelene T. H. Chi, Seokmin Kang & David L. Yaghmourian (2017) Why Students Learn More From Dialogue- Than Monologue-Videos: Analyses of Peer Interactions, Journal of the Learning Sciences, 26:1, 10-50, DOI: 10.1080/10508406.2016.1204546

To link to this article: https://doi.org/10.1080/10508406.2016.1204546

Accepted author version posted online: 23 Jun 2016.
Published online: 11 Aug 2016.

JOURNAL OF THE LEARNING SCIENCES, 26: 10–50, 2017

ISSN: 1050-8406 print / 1532-7809 online
DOI: 10.1080/10508406.2016.1204546

ARTICLES

Why Students Learn More From Dialogue-Than Monologue-Videos: Analyses of Peer Interactions

Michelene T. H. Chi
Mary Lou Fulton Teachers College
Arizona State University

Seokmin Kang
School of Education
University of Wisconsin–Madison

David L. Yaghmourian
Institute for the Science of Teaching and Learning
Arizona State University

In 2 separate studies, we found that college-age students learned more when they collaboratively watched tutorial dialogue-videos than lecture-style monologue-videos. In fact, they can learn as well as the tutees in the dialogue-videos. These results replicate similar findings in the literature showing the advantage of dialogue-videos even when observers watched them individually. However, having the observing students watch collaboratively as dyads provided data to carry out in-depth analyses of their conversations and activities in order to understand why dialogue-videos are superior to monologue-videos. Toward that goal, transcripts of video dialogues and monologues, as well as peer-to-peer conversations of the observing students collected in a prior study, were analyzed using the ICAP (interactive, constructive, active, passive) framework as a lens.

Correspondence should be addressed to Michelene T. H. Chi, Institute for the Science of Teaching and Learning, Arizona State University, P.O. Box 872111, Tempe, AZ 85287-2111. E-mail: Michelene.Chi@asu.edu

Three sets of analyses were carried out. The 1st set focused on the content of the videos in terms of the tutors' and the tutees' moves. The 2nd set focused on the activities and behaviors of the collaboratively observing dyads. The 3rd set focused on the role of the tutees in the dialogue-videos in eliciting *constructive* and *interactive* engagement from the observing students. Our analyses suggest that dialogue-videos naturally elicited more constructive and interactive engagement behaviors from the observers than the monologue-videos, which in turn mediated the observers' own learning.

It has been shown for about three decades now that one-to-one tutoring with either a human tutor or an intelligent tutoring system is the most effective form of instruction for learning (Anderson, Corbett, Koedinger, & Pelletier, 1995; P. A. Cohen, Kulik, & Kulik, 1982; Graesser, Person, & Magliano, 1995; VanLehn et al., 2005). The average effect size is around $d = 0.79$, based on a meta-review (VanLehn, 2011), and the level of learning achieved by tutees is often considered the gold standard (Bloom, 1984). Tutoring is also considered to be personalized in the sense that the instruction is tailored and adapted to the tutee's (the recipient student's) learning needs. Because tutoring is personalized to the tutee, the assumption in the literature has always been that tutees' enhanced learning arises from the tutor's role in delivering individualized and adaptive instruction.

However, in earlier work (Chi, Siler, Jeong, Yamauchi, & Hausmann, 2001), we questioned whether tutees' enhanced learning is derived from the tutor's personalized instruction. Instead, we suggested that the benefit of tutoring could plausibly arise from the many opportunities afforded to the tutees to engage in active learning, such as answering and asking questions, responding to scaffolding prompts, and so forth. Since then, we have developed a new theoretical and evidence-based framework of active learning called *ICAP* (interactive, constructive, active, passive) that supports our conjecture (Chi, 2009; Chi & Wylie, 2014).

The ICAP framework explicitly operationalizes and differentiates what students do to engage with instruction or instructional materials into four kinds of behavioral modes based on the overt activities that students can undertake with the learning materials. The *passive* mode is when students take no overt actions with respect to the instructional materials other than *attending*, such as orienting and listening attentively to an instructor's explanation or lecture or watching the whiteboard while the instructor works out a problem solution. Thus, in contrast to the layman's view that paying attention is required and sufficient for learning, paying attention is only passive in the ICAP framework. The *active* mode is when students undertake any activities that physically *manipulate* the information in the instructional materials without adding any new knowledge, such as underlining text sentences, copying a problem solution, taking verbatim notes that do not provide any new inferences, describing a scenario, reciting a memorized line, and so forth. The next mode is undertaking *constructive* activities, in which students *generate* knowledge beyond what was presented in the instructional materials, such as drawing a diagram, providing an explanation, asking a question, generating a solution, taking notes in their own words (thus producing new knowledge beyond what information was provided in the instruction), and so forth. In the *interactive* mode, two or more peers are *collaborating*, such as coconstructing while dialoguing, by asking and answering each other's questions or elaborating on or challenging each other's comments. *Coconstructing* means that each partner is being constructive/generative, but in a way that is relevant to or builds on the contributions of his or her partners. Sometimes such coconstruction is called *making transactive contributions* (Berkowitz & Gibbs, 1982).

The conjectured underlying knowledge-changing processes for each mode generate the ICAP hypothesis, which predicts that engaging in *Interactive* activities with a partner is more effective for learning than engaging in *Constructive* activities, which in turn is more effective than engaging in *Active* activities, which is superior to being *Passive* (i.e., the I > C > A > P hypothesis). This hypothesis is supported by hundreds of studies existing in the literature when conditional differences are reinterpreted in terms of ICAP modes (see studies cited in Chi, 2009; Chi & Wylie, 2014; Fonseca & Chi, 2011) as well as by our own study (Menekse, Stump, Krause, & Chi, 2013).

According to ICAP, tutoring is beneficial because of the constructive/generative opportunities afforded to the tutees. This prediction was tested in Study 2 in Chi et al. (2001), in which we forbade tutors from giving any adaptive instruction (such as scaffolding and giving feedback). Instead, tutors were only allowed to prompt students to explain using generic or nonpersonalized content-free prompts, such as "What were you thinking there?" Under these generic prompting circumstances, the tutees in the ablated context learned just as well as the tutees who had received personalized tutoring (compare Studies 1 and 2 in Chi et al., 2001). This result confirms ICAP's learner-centered prediction because tutees could be constructive and interactive in both the regular personalized tutoring context and the ablated nonpersonalized context even though the two contexts differed greatly in terms of what the tutors did. Thus, as long as the tutees were engaged at the same ICAP activity level, such as being constructive in the case of responding to tutors' adaptive scaffolding (Study 1 in Chi et al., 2001) or being constructive in the case of responding to tutors' generic prompts (Study 2 in Chi et al., 2001), their learning outcomes should be comparable.

Taking the perspective of students' engagement activities, the ICAP framework gives rise to opportunities to consider alternative forms of instruction that may leverage the advantage of tutoring without the cost, as providing every student with either a human tutor or an intelligent tutoring system is prohibitive. To mitigate the cost and at the same time leverage the benefit of personalized tutoring, we had proposed a novel instructional format in which tutorial dialogues were captured in videos and shown to other learners to watch (Chi, Roy, & Hausmann, 2008). The advantage of this observational format is obvious in that the captured videos can be reused as well as scaled up to be used by hundreds of observing students (Stenning et al., 1999).

In this observational format, in order to receive the same constructive and interactive opportunities that were available to the tutees, the observing learners watched the tutorial videos in dyads to give them opportunities to be collaborative with a peer, and they also had to solve the same problems or do the same worksheets that the tutee in the videos did to give them opportunities to be generative. For example, if the tutee in a video had to answer questions or explain a problem verbally, then we would ask the observers also to write their responses on worksheets. Thus, as per ICAP, we gave the collaboratively observing learners the same constructive and interactive opportunities as available to the tutees, even though the observers were not interacting directly with the tutor, and therefore neither tutor feedback nor tutor questions were personalized to the observing students.

COMPARING TUTORIAL DIALOGUE-VIDEOS WITH LECTURE-STYLE MONOLOGUE-VIDEOS

We have explored this new observational format in two studies (Chi et al., 2008; Muldner, Lam, & Chi, 2014). These two studies differed in two important ways. First, they were carried out in two different and difficult science domains: solving physics problems and understanding the concept of diffusion. Second, one study (Chi et al., 2008) created tutorial dialogue-videos with one expert tutor (defined as an instructor with 30 years of teaching experience), whereas the other study (Muldner et al., 2014) created tutorial dialogue-videos with five nonexpert tutors (who were familiar with the content domain). The similarity in the pattern of results of learning achieved by the college-age observing students suggests that scaling up the production of tutorial dialogue-videos may not require an experienced tutor, thus further reducing the cost of providing individual tutoring on a large scale.

In these two studies, we compared and contrasted the learning advantages of tutorial dialogue-videos to those of lecture-style monologue-videos, as they are the typical format used in a variety of videos for online courses, including massively open online courses (Daradoumis, Bassi, Xhafa, & Cabelle, 2013; Hew & Cheung, 2014). Contrasting these two video formats may shed light on why tutorial dialogue-videos are an effective instructional format.

In both of our studies, the same tutors explained the same concepts and problems in the monologue-videos as in the dialogue-videos in which they tutored, using the same available multimedia presentations. The results from both studies showed that the observers of dialogue-videos learned significantly more than the observers of monologue-videos when the observers watched as dyads. This pattern of results (dialogue-videos being superior to monologue-videos for dyad observers) is consistent with several other studies in the literature in which the two types of videos were watched by solo observers (Craig, Chi, & VanLehn, 2009; Craig, Driscoll, & Gholson, 2004; Driscoll, Craig, Gholson, Hu,

& Graesser, 2003; Muller, Bewes, Sharma, & Reimann, 2008; Muller, Sharma, Eklund, & Reimann, 2007).

In summary, the robust findings for the advantage of tutorial dialogue-videos suggest that we can scale up the benefit of tutoring by creating tutorial dialogue-videos. But in order to know how to create and design tutorial dialogue-videos, we need to understand why they are more effective than lecture-style monologue-videos, even though both types of videos include the same multimedia presentations. Because in our prior studies the observers watched the videos collaboratively, we can analyze the conversations between the dyad peers to shed light on why dialogue-videos are superior to monologue-videos for student learning. The purpose of this article is to present analyses of the data in the Muldner et al. (2014) study. Before explaining the analyses, we provide some details on the method and results of Muldner et al.'s study.

Method of the Prior Study Comparing Dialogue-Videos and Monologue-Videos

Participants

In the Muldner et al. (2014) study, the dialogue-videos were recorded using five adult tutors with a minimal amount of tutoring experience. The tutors were either graduate students in our lab who were trained to understand the concept of diffusion or recruited tutors who had taught diffusion in high school and so were all familiar with the concept of diffusion. Only one tutor had substantial experience tutoring; the other four were inexperienced.

The 10 tutees, as well as the 40 observing students, were university undergraduates who completed the study for a first-year psychology course credit, with an equal number of male and female students assigned to the dialogue- and monologue-video conditions.

Content Covered in the Videos

The content of the videos was solving and explaining the concept of diffusion in the context of seven problem scenarios. Diffusion is a notoriously difficult concept to understand because students misconceive the flow pattern observable in diffusion as similar to other transpositional movement processes, such as water flow (Chi, Roscoe, Slotta, Roy, & Chase, 2012). That is, the flow of water downstream is a linear kind of process, requiring cumulative summing causation, whereas diffusion flow is an emergent kind of process, requiring collective summing causation. Thus, we taught the concept of diffusion in our study because misunderstanding of diffusion is very robust and difficult to remove (Chi et al., 2012), and we wanted to test our replication of the benefit of dialogue-videos on understanding of a concept as difficult as solving physics problems, which was the domain we had used in the earlier Chi et al. (2008) study.

In each video, a tutor explained all of the concepts related to diffusion, such as the concept of concentration, needed to answer the questions embedded in the seven problem scenarios that the tutee had to solve. The seven problem scenarios for this conceptual domain typically consisted of some description of a scenario involving blue dye in containers of water accompanied by a predrawn diagram and displayed on laminated posters. Some scenarios showed both the macrolevel flow pattern of the blue dye and the microlevel movements of the dye and water molecules. For each scenario, two or three questions were asked, such as "Is the behavior of the molecules related to the flow of dye that you see? Explain your answer." Answering such questions usually requires several exchanges.

The videos also had simulations of both the macrolevel and microlevel processes of diffusion that the tutors could control, in terms of stopping and/or repeating segments of the simulations. The purpose of the simulations was to display various aspects of the diffusion process.

Procedure for the Tutors in Creating the Instructional Videos

To prepare the tutors for tutoring, we asked them to read a 2-page college-level text description of molecular diffusion, read the student diffusion workbook containing the seven problem scenarios along with their solutions, familiarize themselves with the macro and micro simulations so that they could use them to answer prompted questions, and listen to a brief tutorial on both effective pedagogical strategy for the tutoring sessions (such as scaffolding or eliciting responses from the tutees) and ineffective pedagogical strategy (such as tutor telling or lecturing to tutees didactically). Pedagogical strategies were based on the findings from our prior work (Chi et al., 2001, 2008). To qualify, tutors had to take a diffusion posttest with a criterion of 80% correct or better. Other than this background preparation and training, and being asked to cover all of the concepts relevant to diffusion needed to answer all of the scenario questions, the tutors were not given any further instruction. That is, neither the dialogues nor the monologues were scripted, and variations in videos were not edited.

For the dialogue-videos, each tutor tutored two college students individually and two middle school students, but we only discuss the results for the college students in order to compare with the results of the Chi et al. (2008) physics study, which was also carried out with college students. Lecture-style monologue-videos were created by having the same five tutors explain the same concepts using the same materials, including simulations. The major difference from the dialogue-videos was the absence of a tutee. The average length of the monologue-videos was 21 min, which was not significantly different from the average length of the dialogue-videos at 25 min (p = .28). Additional details about how the videos were created are described in Muldner et al. (2014).

Procedure for the Students

The 10 tutees and the 40 observing students all read a 2-page diffusion text that provided a general overview of diffusion and then took a pretest consisting of 25 multiple-choice questions. The majority of the questions were taken from Chi et al. (2012). After the pretest, students were assigned using a stratified random sampling procedure either to be a tutee in the tutoring condition or to be a participant in a same-gender pair for either the dialogue- or the monologue-observing conditions.

During the tutoring or the observing phase, students worked on the seven workbook problems either as a tutee (in the tutoring condition) or as an observing student watching collaboratively with a peer either the dialogue- or the monologue-video. Tutees solved the problems by answering the tutor verbally or drawing on the laminated posters, whereas the observing students were supposed to solve or provide answers to the same seven problems in a joint workbook. Although the observing students could not manipulate the simulations embedded in the videos directly, they could indirectly manipulate the simulations by pausing, replaying, or fast-forwarding the videos.

After the video, students took the posttest individually. The posttest consisted of the same 25 multiple-choice questions with four additional questions, for a total of 29 questions. This set of questions consisted of 18 similar questions (i.e., questions that had scenarios similar to one of the seven workbook problems) and 11 transfer questions (i.e., questions with dissimilar scenarios).

The most important feature to note about the procedure is that the observing students in both conditions were asked to be constructive and interactive. That is, not only were students in both dialogue and monologue conditions given the same workbook to complete while observing, but they also were told explicitly to discuss and answer the workbook questions together because each dyad worked to produce a single answer for their joint workbook. Students were also told that they could take as much time as they needed. In addition, students were permitted to control the videos by pausing, forwarding, or rewinding, which allowed them to stop and think about the content. (See Muldner et al., 2014, pp. 72–76, for more detailed description.)

Results for the Pre/Post Gains

We focus on the results of the pre-/posttest gains for the 11 (out of a total of 29) deeper, transfer-type posttest questions, as reported in Muldner et al. (2014), not only because we can compare the results of this study to our prior study in which we also scored correctness on deeper solution steps (Chi et al., 2008), but also because ICAP's predictions can only be tested when deeper learning is assessed. That is, the higher (interactive and constructive) modes of ICAP are predicted to enhance deeper learning, and therefore the benefits of these higher forms of engagement will not be revealed with shallow assessment, making the shallower, nontransfer questions irrelevant for these analyses.

As can be seen in the first two sets of bars in Figure 1, analyzing only the pre-/posttest data for the 11 deeper transfer-type questions revealed that both the tutees and the dialogue-observers learned significantly: tutees, $t(18) = 2.40$, $p = .03$, $d = 1.07$; dialogue-observers, $t(38) = 3.37$, $p < .01$, $d = 1.08$. An analysis of covariance using pretest percentage as the covariate found no conditional differences between dialogue-video observers and tutees, thus replicating our finding that dialogue-observers can learn at the level of the gold standard.

In contrast to the dialogue-observers, the averaged pretest to posttest adjusted gain for the 11 transfer questions of the monologue-observers was not significant ($p = .26$, $d = 0.36$; see the third set of bars in Figure 1). Pairwise comparison of the two observing conditions showed that observers of dialogue-videos learned significantly more than observers of monologue-videos, $t(38) = 2.05$, $p < .05$, $d = 0.07$.

In summary, this prior study (in particular the data reported for the transfer questions in Table 3 of Muldner et al., 2014, and illustrated here in Figure 1) replicated two findings in the literature. First, we replicated our own result (Chi et al., 2008) showing that college students who watch tutorial dialogue-videos can learn as well as the tutees in the dialogue-videos, especially in terms of deeper knowledge. The robustness of this effect was confirmed by the use of five different tutors covering a difficult science concept. Second, we replicated our own finding as well as the findings in the literature that observers learn more when watching tutorial dialogue-videos compared to lecture-style monologue-videos. In fact, the monologue-observers did not learn significantly from pretest to posttest, based on an assessment using transfer-type questions. This pattern of

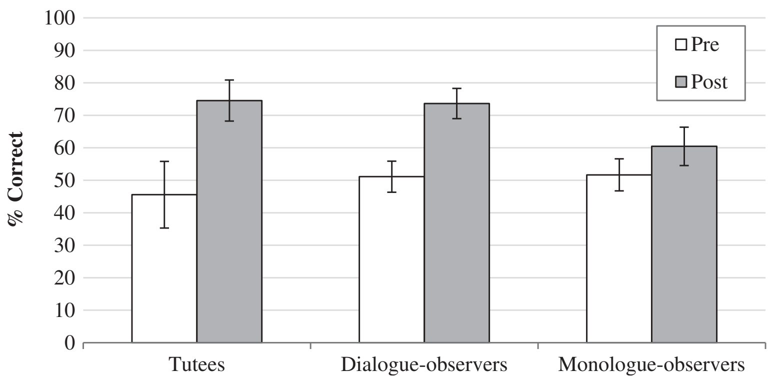

FIGURE 1 Pre- and posttest percent correct for the 11 transfer-type questions. Error bars are ±2 *SE*.

results occurred despite the fact that students in both observing conditions were given opportunities to be constructive and interactive, indicating the efficacy of dialogue-videos.

WHY ARE TUTORIAL DIALOGUES MORE EFFECTIVE INSTRUCTION THAN LECTURE-STYLE MONOLOGUES?

The goal of this study is to carry out three sets of analyses in order to explore and explain why dialogue-videos offer better instruction than monologue-videos for college students. The first set considers the content of the videos, such as tutors' moves and coverage of the key concepts and tutees' moves; these analyses are summarized briefly. The second set considers the engagement behaviors of the observing students; this comprises the bulk of our analyses. The third set of analyses entertains a new hypothesis regarding the role of the tutees in dialogue-videos and provides results in support of this tentative interpretation.

Content of the Videos

The hypothesis explored in this set of analyses was whether tutor or tutee moves in the videos influence the observers' learning. Our analyses segregated content moves that were unique to dialogue-videos from content moves that were common to both dialogue- and monologue-videos. Moreover, we examined the influence of both the tutors' moves, which have traditionally been the only source of analyses for understanding the learning advantage of tutoring, as well as the tutees' moves.

Content Moves Unique to Dialogue-Videos

Dialogue-videos contain many content moves that are absent in monologue-videos. These unique moves can be separated into those uttered by the tutor (e.g., tutor's feedback statements to tutees and tutor's deep questions) and those uttered by the tutee (e.g., incorrect or misconceived statements). These moves are unique to dialogue-videos because a tutee is involved. For example, tutor feedback is typically given when a tutee expresses an incorrect statement; therefore, without tutees, as in monologue-videos, there are not likely to be many tutor feedback statements.

Our analyses considered the tutor moves that prior research has shown to have a direct, beneficial effect on tutees' learning. For example, there is a rich literature showing the importance of the tutor's feedback in facilitating the tutee's learning, especially elaborated feedback, which is superior to simple outcome feedback (Narciss, 2007). In addition, asking the tutee deep questions has also been shown to enhance learning (Craig, Sullins, Witherspoon, & Gloslon, 2006). Because few studies in the literature have examined tutees' moves, we based our choice on analyzing tutee moves that we have found to be effective in our prior studies.

The analyses consisted of coding the transcripts of the tutorial dialogues in each of the dialogue-videos based on the method introduced in Chi (1997). The size of the coding unit used here for each type of move was at the statement level, with each statement consisting of a single idea that could include one or multiple phrases but not necessarily a complete sentence. We used this grain size of analysis because tutors' elaborative feedback and tutees' misconceptions can be more easily identified and interpreted at the statement level.

We then correlated our coded quantities with the normalized learning gains averaged for each dyad. Normalized learning gain was calculated as the pure gain divided by the maximum possible gain, or (posttest score – pretest score)/(1 – pretest score), then averaged for observing pairs to obtain an aggregate normalized gain score for each dyad. Because our sample size was small, it precluded the use of hierarchical linear modeling analyses to account for the nested structure of the data (individuals nested within dyads). Instead, we conducted aggregated analyses as suggested in J. Cohen, Cohen, West, and Aiken (2013) for clustered data by obtaining a mean for each predictor variable and normalized gain score for each dyad.

Throughout this study, because of the massive amount of coding necessary for our deep and intricate analyses, interrater reliability was computed only for those coding rubrics that required subjective interpretation. For others that could be determined more objectively we forwent computing interrater reliability. Interrater reliability was assessed by two independent coders on 20% of the data selected at random. Disagreements were resolved through discussion.

Tutor Moves. Tutors' utterances were coded for two types of moves: elaborative feedback and deep questions. To code for elaborative feedback, we first segmented tutors' speech into statements and identified those that referred to tutees' incorrect or incomplete answers or explanations. We then coded these statements as elaborative feedback if the tutor added explanations beyond saying that the tutee's answer/explanation was incorrect or incomplete. To code for deep questions, we looked at each question tutors asked tutees and determined whether the question (a) required inferences rather than verbatim recall of what had already been discussed and (b) needed more than a yes/no response. Interrater reliability was 98.70% for the coding of deep questions.

Tutee Moves. Tutees' utterances were coded for three types: incorrect or misconceived statements, substantive comments, and questions. To code the first type—whether a tutee made an incorrect statement—we first created a rubric that contained 20 concepts relevant to understanding the concept of diffusion. These 20 concepts were culled from the dialogue- and monologue-videos. (They are shown in Appendix A.) We examined every statement containing a separate idea unit that tutees uttered in the transcript and determined whether it expressed a misconception according to the rubric. For example, in one transcript the tutee said, "They're [meaning the molecules] trying to reach equilibrium." This was coded as an incorrect or misconceived statement related to Concept 6 in Appendix A because molecules do not intentionally try to reach equilibrium. Appendix B gives four examples of tutees' incorrect or misconceived statements and their related concepts. The second type of tutee utterance coded was tutees' substantive comments, which had already been coded in Muldner et al. (2014) at the phrase level. The third type was tutees' questions, which were identified in the current coding at the statement level.

Results. The mean frequencies of each type of tutor and tutee move in the dialogue-videos were correlated with the averaged normalized learning gain per dialogue-observing dyad and are shown in Table 1. Although the literature typically shows that these tutor and tutee moves do affect tutees' learning, there were no significant correlations between these five tutor and tutee moves and the observers' learning.

Content Moves Common to Dialogue- and Monologue-Videos

Very few types of tutor moves are common to dialogue- and monologue-videos. We can think of only two content moves that not only can be expressed readily in both dialogue- and monologue-videos but also provide reasonable hypotheses

TABLE 1
Mean Frequencies (*SD*) of Tutor and Tutee Moves Unique to Dialogue-Videos and Correlations With Dialogue-Observing Dyads' Mean Normalized Gain Scores for All Posttest Questions

		Correlation With Normalized Gain per Dialogue-Observing Dyad (n = 10 Dyads)		
Move	*Dialogue-Videos*	*r*	*p*	r^2
Tutor elaborative feedback	6.80 (3.74)	.33	.36	.11
Tutor deep question	17.80 (6.89)	.09	.81	.01
Tutee incorrect or misconceived statements	3.60 (2.32)	.31	.39	.09
Tutee substantive comments at the phrase level	87.11 (31.79)	−.36	.30	.13
Tutee questions	3.00 (1.78)	.01	.69	.01

that we can generate about them. These two moves are gestures and concept coverage, and the hypotheses we generated about them are discussed next.

Tutor Gestures. The gestures tutors express can be differentiated into two types: iconic and deictic. Iconic gestures bear a close formal relationship to the semantic content of speech. For example, when a tutor describes the path of a molecule, his or her iconic gestures could include tracing the path or indicating the action, shape, and size of objects and other phenomena related to the concept. Deictic gestures, in contrast, are mainly gestures involving pointing at an object or region of space that is given referential value. For example, a deictic tutor gesture might consist of pointing to one molecule in a diagram while saying "This molecule could move to the other side." Thus, iconic gestures provide more congruent and richer information than deictic gestures (Tversky, Jamalian, Segal, Giardino, & Kang, 2014) because they can represent critical components of the diffusion concept, such as the independence of the molecular movement. Based on this difference between the two types of gestures, we hypothesized that iconic gestures could help promote and shape one's understanding of a concept, resulting in enhanced learning from watching videos with a greater number of iconic gestures. Therefore, we were interested to see whether dialogue-videos in fact contain more iconic gestures and, if so, whether they correlated with observing students' learning.

Accordingly, the dialogue- and monologue-videos were meticulously coded for the number of iconic and deictic gestures that were used. Following McNeil (1992), a segment or gesture unit was defined as the period of time between successive rests of the limbs. A hand(s) movement starting from a resting position and returning to the resting position was regarded as one gesture. If the hands did not return to a resting position between two gestures, the boundary was defined by a pause in motion and an obvious change in shape or trajectory. Interrater reliability for the coding of gestures was 91.50% agreement for 20% of the videos.

There was no significant difference in the frequency of tutor gestures between dialogue- and monologue-videos for either iconic or deictic gestures. As we hypothesized, if iconic gestures promoted greater learning, then there should have been a significant correlation between the frequency of iconic gestures and learning for both the monologue- and the dialogue-observers. However, this was not the case for either iconic or deictic gestures (see Table 2). This suggests that the presence of a greater number of iconic gestures in the dialogue-videos could not have accounted for the advantage of dialogue-videos for dialogue-observers.

Concept Coverage. The tutors in the videos were advised to cover the 22 questions embedded in the seven problem scenarios. These 22 questions covered 16 concepts relevant to diffusion. As stated earlier, from all of the dialogue- and

TABLE 2
Mean Frequencies (*SD*) of Tutor Moves Common to Dialogue- and Monologue-Videos and Correlations With Both Dialogue- and Monologue-Observing Dyads' Mean Normalized Gain Scores for All Posttest Questions

	Video Type					*Correlation With Normalized Gain per Dialogue- and Monologue-Observing Dyad (n = 20 Dyads)*		
Move	*Monologue*		*Dialogue*	p	d	r	p	r^2
Tutor gestures								
Iconic	57.70 (26.42)	=	66.60 (47.32)	.61	.23	.16	.51	.03
Deictic	247.20 (115.04)	=	175.30 (105.49)	.16	.65	.03	.90	<.01
Tutor covered concepts	14.50 (2.42)	>	10.90 (2.88)	<.01	1.35	.03	.90	<.01

monologue-videos, we culled all of the unique concepts relevant to diffusion that were expressed by either the tutors or the tutees and found 20 such concepts (shown in Appendix A). The four concepts that were not queried were Concepts 2, 13, 14, and 18 in Appendix A. Out of these 20 concepts, we coded how many the tutors expressed in dialogue- and monologue-videos. Interrater reliability was 81.48% agreement for identifying when a concept was expressed and 98.14% for the identity of a concept (out of the 20 in Appendix A) for 20% of the video transcripts. Somewhat surprising is that the tutors expressed significantly fewer relevant concepts in the dialogue-videos than in the monologue-videos (10.90 vs. 14.50 concepts, respectively), $F(1, 18) = 9.16$, $p < .01$ (see Table 2).

Despite the significantly fewer number of concepts covered by the tutors, there was again no significant correlation between the frequency of concepts tutors covered and the average normalized learning gain per dyad for both dialogue- and monologue-observing dyads (see Table 2, last three columns). This lack of correlation may be due to the tutees in the dialogue-videos initiating an average of 3.6 relevant concepts. Thus, overall adding the tutee-initiated concepts brings the number of concepts mentioned in the dialogue-videos to 14.50, which is equivalent to the number mentioned in the monologue-videos.

The fact that on average tutors in dialogue-videos covered fewer of the relevant concepts than tutors in the monologue-videos seems logical because tutors engaged in a dialogue with a tutee must tailor their explanations of a concept to the tutee's understanding, thus spending more time on a given concept as needed. We can verify this personalization interpretation in the following way: Because each instructor tutored two tutees in the dialogue-videos and explained lecture style twice in the monologue-videos, we can calculate the differences in the tutorial length between the two dialogue-videos compared to the two monologue-videos with one measurement for each instructor. We expect greater variability within tutors in the length of their two dialogue-videos, on average a 7.8-min difference between the two dialogue-videos for two tutees, whereas the same instructor's/tutor's monologue-videos seem more uniform in length, with a 1.4-min difference. The within-tutor variability in length for each pair of a tutor's dialogue-videos suggests that the tutors must have adapted to individual differences in their tutees' understanding, which could have included spending more time on a concept for tutees requiring additional help. This within-tutor variability in the length of tutoring further supports our conjecture proposed earlier that this personalization, tailored to the tutees (not the observers), could have reduced the amount of time available for covering all of the concepts in the time allotted.

The overall pattern of our findings, despite the painstaking hours of coding, is that none of these seven factors (shown in Tables 1 and 2) correlated with the observing students' learning outcomes. For example, the number of tutors' deep questions, on average 17.80 for the dialogue-videos, did not correlate significantly with dialogue-observers' learning ($r = .09$, $p = .81$). Thus, we tentatively conclude that overall, tutors' and tutees' moves, tutors' gestures, as well as the completeness of content coverage did not contribute directly toward the dialogue-observers' greater learning, which suggests that the advantage of the dialogue-videos is not in the hands of what the tutors and the tutees did per se.

Observers' Own Engagement Behaviors

The ICAP hypothesis predicts that how well students learn depends largely on how actively they engage with each other and with the instructional materials, which in this case were the videos. According to ICAP, the same video lecture can cause greater learning if students engage *constructively* or *interactively* while watching (e.g., solving the problems on the worksheet and collaborating with their partner) rather than *actively* (e.g., copying the problem solutions that are presented in the video). Accordingly, instead of analyzing the influence of the content of the videos (as expressed in the feedback, the deep questions, the gestures, etc.) as in the first set of analysis, in this second set of analyses we explored whether there were differences in the way in which the dialogue- and monologue-observers interacted with each other and with the videos. We then correlated these factors with the observing students' learning.

We had two sets of data to analyze to capture observing students' collaborative behavior: their conversations and their joint workbook. For clarity, we refer to the observing students' peer-to-peer dialogues as *conversations* to disambiguate the term from the tutor–tutee tutorial dialogues in the dialogue-videos. We first explored the number of *constructions* and *interactions* the dyad observers engaged in for each type of video. Similar to the previous analyses, we did not determine interrater reliability for analyses that required an objective recomputing of coding carried out previously that had already reported interrater reliability. Such recomputation included summing previously coded units, such as the number of substantive comments within problem scenarios, for which the problem scenarios could be easily identified. However, for other boundaries that were more subjective, such as *interaction episodes*, we determined interrater reliability (see below).

Analyses of Observing Dyads' Conversations

The observing dyads' conversations while watching the monologue- and dialogue-videos had been transcribed for the analyses in the previous study (Muldner et al., 2014), segmented into statements, and coded.

Constructive: Substantive Comments. Phrases were coded as substantive comments if they pertained to an idea relevant to the concepts being taught, regardless of whether the comments were correct or not. The following are examples of two substantive/relevant comments at the phrase level: "I think it would be in all directions," (referring to the flow of dye) and "because they are always moving" (referring to the molecules). Non-substantive comments were either comments irrelevant to the topic of diffusion or comments such as "Okay," "So," or "I agree."

In order to be *constructive*, students had to utter substantive comments that contain ideas that went beyond what was already expressed in the instructional materials or instruction. However, here we simply take substantive comments as evidence of construction because such coding was already carried out in Muldner et al. (2014) to obtain an overall mean per person, and it was reported in that article that each member of the dyad dialogue-observers generated overall a greater number of substantive comments (45.35 total per person) than each member of the dyad monologue-observers (28.21), and this difference was significant with a large effect size ($p = .03$, $d = 0.73$). Interrater reliability for this original coding is reported in Muldner et al. ($\kappa = .88$ for 20% of the transcripts). These data are shown in Table 3, row 1. Thus, we already know from prior analysis that dialogue-observers were more constructive than monologue-observers.

Interactive: Frequency and Richness. Dyads' conversations could be further analyzed to see how *interactive* the dyad observers were. Using the same transcripts of the dyads' conversations, we first segmented them into episodes. An *episode* was a portion of the observers' multiturn conversation on the same topic and line of thought that included at least one substantive comment from either member of the dyad, with substantive comments as defined in the preceding analysis. Interrater reliability for determining the boundaries of episodes was 86.67% agreement for 20% of the observer transcripts.

TABLE 3

Mean Frequencies (*SD*) of Observing Dyads' Engagement Behaviors per Video While Watching Monologue- and Dialogue-Videos and Correlations With Their Mean Normalized Gain Scores for All Posttest Questions

Behavior	*Video Type*					*Correlation With Normalized Gain per Dialogue- and Monologue-Observing Dyad (n = 20 Dyads)*		
	Monologue		*Dialogue*	p	d	r	p	r^2
Conversations								
Frequency of substantive comments per person	28.21 (14.18)	<	45.35 (26.12)	.03	.73	.50	.03	.25
Frequency of interaction episodes per video	9.30 (4.73)	=	11.30 (5.39)	.22	.39	.41	.07	.17
Substantive comments per interaction episode	4.98 (4.71)	<	6.10 (4.14)	<.01	.43			
Coconstructive turns per interaction episode	2.18 (2.01)	<	2.47 (2.15)	.01	.14			
Workbooks								
Solved workbook problems	10.90 (7.16)	<	16.70 (4.90)	<.05	.97	.24	.32	.06
Jointly solved workbook problems	5.30 (4.14)	<	10.50 (5.42)	.03	.96	.38	.10	.15

Each episode was then further identified as an *interaction episode* if both observers contributed at least one substantive comment. For example, in Appendix C, Episode 1, on the topic of nonintentional properties of molecules, was defined as an interaction episode because Speaker F contributed substantive comments in lines 1 and 3 and Speaker M contributed a substantive comment in line 2. However, Episode 6, on the topic of the observable pattern of diffusion, is not an interaction episode because only Speaker K made substantive comments in lines 1, 2, and 4, and Speaker Z's comments in lines 3, 5, and 7 were not substantive. Using this coding rubric, we found that monologue-observers participated on average in 9.30 interaction episodes per video, whereas dialogue-observers participated in 11.30 interaction episodes per video (percent agreement was 86.67%; see Table 3, row 2).

Although this difference in the mean number of interaction episodes did not reach significance between the dialogue- and monologue-observers, there was a lack of overlap in the frequencies of interaction episodes for the majority of the monologue- and dialogue-observers. That is, the majority (seven of the 10 pairs) of the monologue-observers interacted between four and nine episodes per video, whereas the majority (seven of the 10 pairs) of the dialogue-observers interacted between 12 and 17 episodes per video. This suggests that perhaps a finer grained analysis of interaction quality was needed to reveal differences between dialogue- and monologue-observers. Accordingly, the richness of the interactions was captured in the following more fine-grained ways that reflect their *coconstructive* quality.

First, we hypothesized that an interaction episode that contained more substantive comments would be richer in that such an episode would contain more content-relevant information than one with fewer substantive comments. We found that on average dialogue-observers generated 6.10 substantive comments per interaction episode versus 4.98 substantive comments per interaction episode for the monologue-observers, a significant difference, $\chi^2(1, N = 205) = 3.55$, $p < .01$.[1] Thus, the interaction episodes of dialogue-observers contained significantly more substantive comments than the interaction episodes of the monologue-observers, which shows that the dialogue-observers were more constructive overall in their interactions.

Second, from the ICAP perspective, we also hypothesized that an interaction episode that contained more coconstructive turns would be richer than an interaction episode with fewer coconstructive turns. This hypothesis was based on the idea that coconstructing generally means that each partner builds on, refines, or challenges the ideas of the other partner (Chi & Wylie, 2014). Thus, we coded

[1]Poisson regression was used to model count variables with the assumption that the conditional means equaled the conditional variances.

the frequency of coconstruction as a turn in which both speakers contributed substantive comments consecutively.

A *turn* is usually defined in the literature merely as a change in speakers. However, a *coconstructive turn* is defined here as a change in speaker that contains substantive contributions from both speakers. For example, in Appendix C in Episode 1, Speaker F provided a substantive comment at the start of turn 1 (indicated by the brackets). Immediately afterward, Speaker M provided a substantive comment. Thus, a turn in which both speakers contribute substantive comments is one coconstructive turn. In Episode 3, Speaker M provided a substantive comment at the start of the episode. However, Speaker F did not provide a substantive comment in response in line 2. Speaker M continued to make substantive comments in lines 3 and 4. Speaker F finally made a substantive comment in response in line 5. Therefore, Episode 3 has just one coconstructive turn because a coconstructive turn requires that a substantive response be made to a substantive comment. Thus, our definition of a coconstructive turn is more restrictive than the definition in the literature, in which a turn consists merely of a change in speaker, because our definition captures the coconstructive nature of interactions.

Per interaction episode, the number of coconstructive turns taken by the dialogue-observers exceeded the number of coconstructive turns taken by monologue-observers (2.47 vs. 2.18 per interaction, respectively). Categorizing fewer than three turns as low and greater than three turns as high, a nonparametric test showed that the difference was significant, $\chi^2(1, N = 206) = 6.71, p = .01$. Thus, when dialogue-observers interacted, there were more back-and-forth substantive coconstructive exchanges within each episode than among monologue-observers.

In summary, although the overall frequencies of interactions did not differ significantly between monologue- and dialogue-observers, closer inspections of the quality of the interactions revealed that dialogue-observers interacted in a richer way in that they generated more substantive comments and engaged in a greater number of coconstructive turns than the monologue-observers. Overall this pattern of results from the analyses of the conversations indicates that the interactions of the dialogue-observers were more constructive and coconstructive than the interactions of the monologue-observers, and such engagement behaviors could have contributed to the greater learning of the dialogue-observers.

Analyses of the Workbook

The workbook contained seven problem scenarios, with each scenario asking approximately three questions, such as “What direction will the ink molecules be going?”, for a total of 22 questions per workbook. Thus, the data to be analyzed consisted of answers and explanations for these 22 questions. Each dyad had one joint workbook that was identical to the one the tutees worked on.

Constructive Versus Active: Solve Versus Copy. When dyads of observing students work on their joint problem workbook, they can either simply copy what the tutor or the tutee in the video wrote or solve each activity on their own. Based on the ICAP framework, *solving* a problem is more likely to be a generative/constructive activity than *copying* a problem, which is more likely to be a manipulative/active activity; thus, solving should foster more learning than copying. Therefore, we hypothesized that the dialogue-observers should be solving more than copying as well as solving more than the monologue-observers.

A problem solution/explanation was coded as copied when observing students wrote exactly the same words right after a tutor or a tutee gave the answer in the instructional video. A problem solution/explanation was dubbed solved when the observers wrote down their answer after discussing it with each other (coders could overhear from the video when the members of the dyads were discussing with each other) and the answers were paraphrased in their own words or used different terms than what was presented by the tutor or tutee. Because this coding of copied versus solved required (a) examining the worksheet answers along with viewing the videos to see what was written on the laminated posters and (b) listening to the dyads' conversations, 20% of the dyads' workbook questions were coded by a second coder to assess interrater reliability (percent agreement = 94.32%).

Overall the dialogue-observers solved their workbook problems significantly more often than they copied the answers ($M = 16.70$ for solve vs. 5.30 for copy out of 22 questions), $F(1, 18) = 29.19$, $p < .01$. In contrast, the monologue-observers solved about equally as often as they copied (10.9 times for solve and 11.1 times for copy). Not only did the dialogue-observers solve more than copy, they also solved a significantly higher proportion of their 22 workbook questions than did the monologue-observers, $F(1, 18) = 4.47$, $p < .05$ (see Table 3, row 5). This is consistent with ICAP's prediction that dialogue-observers learned more than the monologue-observers because the dialogue-observers solved, a constructive behavior, more than they copied, an active behavior.

Interactive: Jointly Solve. Whether dyads were coded as having solved the problems rather than copied them was determined by whether the problem answers were identical to the answers given in the video and whether there were some discussion between the dyad members about the answers. However, we can further determine whether a problem was solved jointly by identifying whether both partners contributed substantive comments to the answers. If both partners contributed substantively, then they collaborated to solve the problems. This coding revealed that dialogue-observers jointly solved twice as often (10.50 out of 22 questions) as the monologue-observers jointly solved (5.30 out of 22 questions), and the difference was significant, $F(1, 18) = 5.81$, $p = .03$ (see Table 3, row 6).

In summary, for both sets of data—the dyads' conversations and their workbooks—we found that dialogue-observers were more constructive and interactive than monologue-observers. More constructive meant that they either generated more substantive comments or undertook a more generative task, such as solving versus copying. More interactive meant that their multi-turn interaction episodes had more substantive comments and more coconstructive turns and that their problems/questions were more frequently jointly solved. Thus, overall the dialogue-observers engaged in higher modes of active learning than the monologue-observers, which accounts for their improved learning as assessed by the 11 deeper transfer-type questions (as shown in Figure 1).

Correlations With Learning

Although it is tempting to conclude that the dialogue-observers must have learned more because they engaged more actively (i.e., they were more constructive and they interacted in a richer fashion, as the preceding set of analyses showed), we need to see whether there is in fact a relationship between the frequencies of their active modes of engagement and the learning of both monologue- and dialogue-observers, as an active engagement activity should benefit all observers. The ICAP hypothesis basically predicts that one learns more in one mode (such as the constructive mode) than another mode (such as the active mode). Although it is not explicitly stated, the ICAP hypothesis implies that the more frequently one engages in a specific mode of activity, whether constructive or interactive, the more one learns, as indicated by the self-explanation effect (Chi, de Leeuw, Chiu, & LaVancher, 1994). Hence, our available data provide an opportunity to test whether greater frequency of a specific ICAP mode of engagement activity fosters greater learning.

Because engaging in active learning should be beneficial to both monologue- and dialogue-observers, we correlated the frequencies of both the monologue- and the dialogue-observers' substantive comments, interaction episodes, and solved and jointly solved frequencies with their normalized gain scores for all of the posttest questions, averaged for each dyad. All correlations were with each dyad's average normalized learning gains because there are dependencies between the peers within a dyad. We found a significant correlation for substantive comments ($r = .50$, $p = .03$), a marginal correlation for interaction episodes ($r = .41$, $p = .07$), a marginal correlation for jointly solved ($r = .38$, $p = .10$), but no significant correlation for the frequency of problems solved ($r = .24$, $p = .32$; see Table 3, last two columns). Thus, there was some suggestion with trending effects that the more frequently students are constructive and interactive, the more they learn, with the exception of solving.

Because solving is a constructive activity, we expected it to also correlate with all of the observers' learning, even if only marginally. We explain this lack of correlation for solve across both the monologue- and dialogue-observers by a curious finding that shows a different pattern in the learning gains of the partners within dyads for monologue- and dialogue-observers as they solved. Basically the dialogue-observing dyads had an equivalent amount of learning gains between them, whereas the monologue-observing dyads had discrepant amounts of learning gains. This disparity in the pattern of learning gains between the partners of dialogue-observing dyads and monologue-observing dyads could have accounted for the lack of correlation in the number of solve episodes in Table 3, as the averaged gain for the monologue dyads was lower. This finding further suggests that the dialogue-observers are benefitting from working together, whereas one partner of the monologue-observing dyads is dominating and solving the workbook problems alone.

In summary, in this second set of analyses, we coded the observing dyads' constructive and interactive behaviors and found a systematic pattern of results in that the dialogue-observers were more constructive and interactive than the monologue-observers. In addition, consistent with ICAP's prediction, there was a trending effect (with the exception of solve) showing the benefit of being more constructive and more interactive on the learning gains for both monologue- and dialogue-observers, which suggests that the greater frequencies with which the dialogue-observers participated in constructive and interactive behaviors could have enhanced their learning from dialogue-videos.

The Role of the Tutee

The preceding analyses suggest that compared to monologue-videos, dialogue-videos tended to elicit higher modes of active engagement from the observing students, with behaviors such as generating more substantive comments, solving problems more than copying solutions, and interacting in a richer and more coconstructive fashion. Because these elicited behaviors are either constructive or interactive, it makes sense, from an ICAP perspective, that they mediated the observers' greater and deeper learning from the dialogue-videos than from the monologue-videos, and the correlations show a trend, even for such a small sample. But the question remains: Why do dialogue-videos elicit more generative and collaborative behaviors from the observing students? The dialogue-videos also seemed to trigger more active engagement from the observers naturally, in that there were no explicit or intentional prompts embedded in the dialogue-videos. For example, although the tutors asked many questions, the questions were not directed at the observing students, and likewise the observing students were not being generative as the result of trying to answer the tutors' questions. This is consistent with finding no significant correlation between the tutors' deep questions and the observers' learning. Thus, the observing students seemed inclined to be more constructive and interactive naturally, that is, without any explicit elicitations from the tutors.

We know that the pedagogy of the dialogue-videos and the monologue-videos is vastly different, mainly in terms of the presence or absence of a tutee. In this section, we forward a new hypothesis for why observing students want to engage more actively when watching dialogue-videos, and our hypothesis focuses on the role of the tutees rather than the tutors. This hypothesis focuses on answering the questions of how important the role of the tutee is in the dialogue-videos and why.

The role-of-the-tutee hypothesis can be decomposed into three corollary hypotheses that guided our analyses in assessing the role of the tutees: First, we analyzed the observing students' targets of referral; second, we analyzed the extent to which tutees served as a model of learning; and third, we analyzed the more specific role of conflict episodes, or episodes in which tutees' errors were followed by tutor feedback.

The Target of Referral

We propose that tutees play a significant role in dialogue-videos because they offer the benefit of social presence, even though they are not physically present (i.e., they are only in the videos). Although the majority of the extensive literature on the benefit of social presence refers to direct physical social presence (Biocca, Harms, & Burgoon, 2003), there is some emerging evidence suggesting that indirect social presence also plays an important role. In a study by Okita, Bailenson, and Schwartz (2007), adult participants interacted with a character in an immersive virtual reality. The participants were told either that the character was a computer agent or that it was another person in another room. When the responses from the character were held constant, the participants learned more (about the biological mechanisms of fever) when they thought the character was a real person. The amount of their learning was also correlated with higher levels of arousal as measured by skin conductance. Although Okita et al.'s results show that interacting with an assumed remote presence of a social being benefits learning more so than interacting with a computer agent, their results also suggest that perhaps watching social interactions in dialogue-videos may motivate our observing students' learning more so than watching monologue-videos in which there is no social interaction.

To test whether the visual recorded presence (but not physical presence) of a tutee matters for the dialogue-observers, we can examine whether their conversations referred to the tutees at all and, if so, whether they referred more to the tutees or more to the tutors. A logical hypothesis, based on decades of tutoring research that focused on tutors' moves (P. A. Cohen et al., 1982), is to expect the observers to pay more attention to the tutors because the tutors are the experts and the instructors.

We first tested this expectation in Chi (2013) by reanalyzing the conversations between the observing dyads of the Chi et al. (2008) study in the following way. We segmented the dyad observers' conversation into episodes, with each episode addressing only one concept or one problem-solving step. We then narrowed down and focused only on episodes in which the dyad observers were trying to resolve misunderstandings and make sense of the content in a given segment of the video (as opposed to episodes in which they were not referring to the video). Within each of these resolving episodes, we coded whether the paired observers' discussion referred to what the tutee did and said or what the tutor did and said as well as referrals to the whiteboard or referrals to both the tutor and the tutee. The analysis of interest is to contrast referrals to the tutor versus the tutee. We were surprised to find that the dyad observers referred significantly more often to what the tutees said (6.8 episodes per problem) than what the tutor said (1.8 episodes per problem; $p < .02$). This result was counterintuitive, as we expected more references to what the tutor said, consistent with the common assumption stated previously and the fact that tutees themselves do pay attention to (and presumably learn from) what tutors say.

In the current data set, we analyzed both the monologue- and the dialogue-observing students' referrals in the same way. Using the episodes identified earlier (based on the content of the dyads discussing the same topic or line of reasoning, with examples in Appendix C), we coded only those episodes that referred to the video. This allowed us to see whether the target of referrals was the tutor or the tutee.

Figure 2 shows the results. Because the monologue-observers could only refer to the tutor, summing the left two bars shows that on average they referred to the tutor a total of 12.70 times, which is serendipitously identical to the total number of referrals (12.70) the dialogue-observers made to both the tutor and the tutee (summing all four bars on the right side of Figure 2). This equivalence in the total number of referrals suggests that both groups of observers paid attention to the videos to the same extent. However, the dialogue-observers referred to what the tutee said ($M = 8.70$, $SD = 5.03$, summing columns 3 and 4 of dialogue-observers) significantly more often than they referred to what the tutor said ($M = 4.00$, $SD = 2.49$, summing columns 1 and 2 of dialogue-observers; see Figure 2), $F(1, 18) = 7.00$, $p = .02$. This confirms our prior finding reported previously (in Chi, 2013) for the data in the Chi et al. (2008) study. This means that monologue-observers referred to what the tutor said significantly more often (12.70 referrals per video) than the dialogue-observers (4.00 referrals per video),

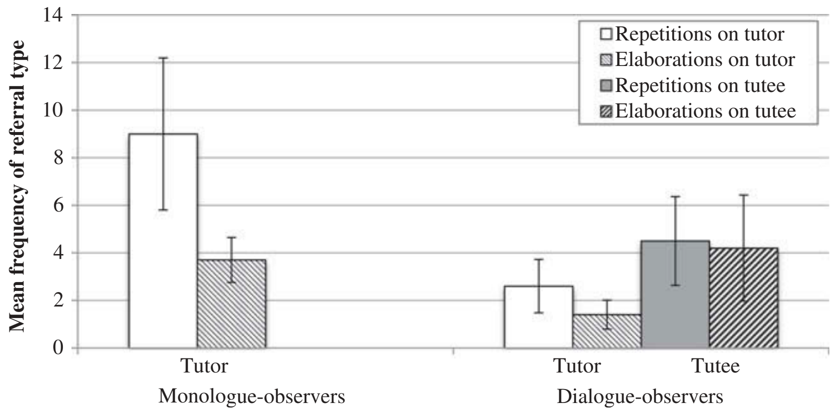

FIGURE 2 Frequency and type of referrals to tutors and tutees. Error bars are ±2 *SE*.

$F(1, 18) = 20.51$, $p < .01$ (see the first two sets of bars in Figure 2), and yet referring more to the tutors was not beneficial to the monologue-observers' learning.

The bars in Figure 2 also show whether the referrals *repeated* (solid bars) or *elaborated* (hatched bars) what was being said by either the tutor or the tutee. Elaborated means that additional ideas and inferences were generated that extended beyond what the tutor or tutee said. The first set of bars shows that the monologue-observers repeated what the tutors said (9.00 repetitions per video) significantly more often than they elaborated on what the tutors said (3.7 elaborations per video), $F(1, 18) = 10.108$, $p = .005$. This shows that the monologue-observers were more active than constructive, because repeating only manipulates the tutors' ideas (and thus is an active form of engagement), whereas elaborating the tutors' ideas adds or infers more knowledge, according to ICAP. This may explain why the monologue-observers did not benefit much from the tutors' lectures. Although the dialogue-observers also did not elaborate much on what the tutor said (on average 1.4 elaborations), they did elaborate significantly more often on what the tutee said (4.2 vs. 1.4), $F(1, 18) = 5.880$, $p = .026$.

In sum, although overall both the monologue- and the dialogue-observers referred to the videos to the same extent, at least in this laboratory study, the targets of the observers' referrals differed substantially. Having a tutee present in dialogue-videos allowed the dialogue-observers to refer to and elaborate on what the tutees said more often than what the tutors said. It is curious that referring to what the tutees said, which may be more incorrect than what the tutors said, could be helpful to the dialogue-observers. The novel hypothesis we forward in the next section may explain this unexpected referral finding.

Imitating Learning Skills: A Model of Learning

Why does the social presence of the tutee provide an advantage? Social presence is merely a construct that labels the phenomenon, suggesting that having a tutee present is important. Our preceding analyses confirm that when a tutee was available, observers referred to and elaborated on what the tutees said more so than the tutors. However, neither the referral results nor the construct of social presence explains why or how having a tutee is beneficial; that is, what do the observers learn from the tutees? The benefit of having a tutee is not direct, in the sense that there were no significant correlations between tutees' moves and dialogue-observers' learning, as shown in Table 1. With this data set, we can test two hypotheses for why a tutee's presence may mediate or facilitate dialogue-observers' learning: One hypothesis is that a tutee provides a model of learning, and the second hypothesis (to be addressed in the next section) is that attempts to resolve a tutee's errors may also help observers' learning.

To claim that the availability of a tutee in dialogue-videos affected observers' learning, we must show a more direct relationship between the tutee's behaviors and the dialogue-observers' behavior. One relationship may be that the observing students perceive the tutee as a model of learning skills; therefore, they might mimic how tutees learn, such as asking many questions. Two studies in the literature have shown that college-age students can mimic and learn learning skills through observation. For example, Craig, Gholson, Ventura, Graesser, and Tutoring Research Group (2000) found that observing students can learn to ask questions by watching an animated agent ask questions. Similarly, Rummel and Spada (2005) showed that observing students can learn the skill of collaborating by watching others collaborate. In short, these studies provide evidence that students can learn such constructive (e.g., asking questions) and interactive (e.g., collaborating) learning skills by watching others display these skills.

Observing learning skills displayed by others may have more of an impact than we suspect because students rarely see good learning skills displayed in conventional classrooms. For example, a student in a conventional classroom asks questions at the rate of only 0.11 questions per hour (Graesser et al., 1995). Therefore, it may be a fairly novel experience for observing students to see tutees ask questions or give substantive comments; consequently, paying attention to the tutees may facilitate learning these skills.

To see whether observers mimic tutees' constructive and interactive behaviors, we can correlate the frequency of the tutees' constructive behavior with the observers' constructive behaviors. One constructive behavior is the frequency with which the tutees asked questions, which can be counted from the videos, as well as the frequency with which the dialogue-observers asked each other questions. When the number of questions tutees asked per problem scenario (0.43) was correlated with the number of questions the observing students asked each other (1.11) when they watched the corresponding dialogue-videos,

TABLE 4
Mean Frequencies (*SD*) and Correlations Between the Constructive Behaviors of Tutees and Dialogue-Observers

			Correlation		
Behavior	*Tutee*	*Observing Students*	r	p	r^2
Questions per problem scenario (n = 70)	0.43 (0.71)	1.11 (1.34)	.30	.01	.09
Substantive comments per problem scenario (n = 63[a])	10.54 (7.38)	12.71 (10.77)	.34	<.01	.12

Note. The frequency of substantive comments for observing students is the mean total per dyad.
[a]With outlier removed.

we were surprised to find a strong relationship (r = .30, p = .01; see Table 4). This suggests that dialogue-observers may have noticed and imitated tutees' question-asking skill (see Table 4, row 1).

Another constructive behavior that tutees express is generating substantive comments rather than just saying "Okay" or "Um yeah," which are not substantive comments. Using the coding previously done in Muldner et al. (2014), we correlated the number of substantive comments the tutees generated per problem scenario (10.54, as shown in Table 4) with the number of substantive comments the observing students generated per pair per problem scenario (12.71). Again, this correlation was also significant (r = .34, p < .01, see Table 4, row 2).

In short, the two correlation results shown in Table 4 provide some evidence that the observing students may have been imitating the constructive learning skills (i.e., asking questions and generating substantive comments) of the model tutees. This would explain the greater learning benefits that come from observing dialogue-videos because observers adopt the learning skills displayed by the tutees, which causes them to be more constructive and interactive, as described earlier.[2]

Resolving Conflicts or Reacting to a Struggling Tutee

Tutees provide errorful behavior in the sense that they can give incorrect explanations, say incorrect statements, and so forth. Tutees' incorrect statements are typically followed by tutors' corrective and/or elaborative feedback statements. For example, in Chi et al.'s (2001) tutoring data, we found an overall one-

[2]Ideally a mediation model analysis would be used to further investigate whether observers' behaviors analyzed in these correlations influenced the effect of tutee moves on observers' learning. However, the sample size was not large enough for such a model.

to-one correspondence between tutees' expression of incorrect statements followed by tutors' corrections or feedback. To confirm that this one-to-one correspondence of error followed by feedback also occurred in our current data, we examined the 27 out of 70 total problem scenarios (seven scenarios per video for 10 dialogue-videos) that contained tutee incorrect statements. Of these 27 problem scenarios in which incorrect statements were expressed, tutors gave feedback for all of them (even if it took more than one statement). Thus, for this data set, tutees' incorrect statements are also tightly linked with tutors' feedback. Therefore, instead of separating these incorrect statements and tutor feedback as two separate moves, it makes more sense to refer to such coupled moves at a larger grain size of a conflict or a conflict episode. In this section, we consider our second hypothesis addressing why the conflict episodes involving tutees might enhance dialogue-observers' learning.

There is some evidence in the literature that watching conflicts does facilitate learning. For example, Schunk, Hanson, and Cox (1987) showed videos of subtraction problem solving by either a struggling (coping) student or a mastery student with an instructor, and children learned the subtraction skill more effectively from observing the coping student. Similarly, Monaghan and Stenning (1998, Study 1) also showed that adult students who watched a video of a model student struggling to solve syllogism problems learned slightly more (albeit not significantly more, $p > .05$) than students who watched a model sailing through the problem solving.

Why conflict episodes in tutorial dialogues might enhance observers' learning has been explored in the literature by one set of investigators to explain why learning from observing tutorial dialogues is better than learning from observing lecture-style monologues. The hypothesis offered in the literature, as proposed by Muller et al. (2007, 2008), is that dialogue-observers are exposed to incorrect tutee statements that are followed by their corrective refutations, and these feedback inputs provide rich information from which observing students can learn. Thus, Muller et al.'s hypothesis is that the content of these incorrect statements, along with their refutations, provides the correct information that facilitates the observers' learning. However, our results show that there were no direct significant correlations either between tutees' incorrect statements and observing students' learning (see Table 1, row 3) or between the tutors' feedback and observing students' learning (see Table 1, row 1). Because the two moves are linked, it makes sense that the same pattern of nonsignificant correlational results was found for both moves. This pattern of correlation was also found in our 2008 data (see Table 3; Chi et al., 2008). That is, for both sets of data, there were no direct relationships between either the tutees' incorrect statements per se, or the tutors' refutation feedback to these incorrect statements per se, and observers' learning, which suggests that the content of these incorrect and refutation statements is not the source from which observing students learned.

We offer instead a novel hypothesis for why dialogue-observers learn from conflict episodes: We surmise that conflict episodes could help the observers learn not because they provide refutation information but because the displayed conflicts motivate the observers, not only extrinsically for the practical reason of getting a correct answer (given that the tutee obviously has gotten it wrong in the conflict episodes) but intrinsically to try harder and put in more effort. We capture trying harder in terms of evidence engaging in more effortful constructive and interactive activities. Our interpretation that observers might try harder when tutees make errors and receive corrective feedback is supported by work on cognitive conflict in the context of learning science concepts. Chan, Burtis, and Bereiter (1997) asked ninth- and 12th-grade students to think aloud or discuss with their peers eight scientifically valid statements that conflicted with their existing beliefs to varying degrees. Students' verbalizations were coded to see whether they assimilated new information with what they knew or treated the new information as something problematic that needed to be explained. In general, when students experienced conflicts, they tended to engage in a knowledge-building type of constructive activity when trying to explain the conflict. Although these conflicts caused the person who expressed incorrect statements to be more constructive, the same reaction may be present for observers of conflicts.

Although our data already show that dialogue-observers were more likely to solve the problems themselves rather than rely on and copy what the tutees did (see Table 3), this result is consistent with both the extrinsic motivation interpretation (getting the right answer) as well as the intrinsic motivation interpretation (trying harder). Thus, we present three additional sets of evidence showing that dialogue-observers were more constructive and interactive in ways other than in their frequency of solve versus copy.

First, if our hypothesis is true that seeing conflicts may cause observers to be more constructive and interactive, then we should see significant correlations between the number of conflicts and observers' frequencies of (a) generating substantive comments, (b) solving, (c) and interacting. Because we had already coded tutee incorrect statements (M = 3.60 per video; see Table 1) and tutor feedback statements (M = 6.80 per video; see Table 1) independently, rather than recode each tutee incorrect statement followed by tutor feedback at the grain size of a conflict episode, we just correlated both the tutee statements and the tutor feedback statements with observers' constructive and interactive activities separately, expecting the same pattern of correlations because they are linked. It is not surprising that all of the correlations between tutees' incorrect statements and tutors' feedback statements and observers' constructive and interactive behaviors were significant (see Table 5). This suggests that seeing a tutee struggle in conflict episodes encouraged the observing students to try harder (in terms of generating more substantive comments, solving more, and interacting more), which could have mediated learning, even though there

TABLE 5
Correlations Between Conflict Episodes, Coded Separately as the Number of Tutee Incorrect Statements and the Number of Tutor Feedback Statements, and the Mean Frequencies per Dyad of Dialogue-Observers' Constructive and Interactive Behaviors

Tutee/Tutor Conflict Moves	*Observers' Const/Interact Behaviors*	*r*	*p*	r^2
Tutee incorrect statement	Substantive comments	.37	<.01	.14
	Solve frequency	.31	.01	.09
	Interaction episodes	.44	<.01	.19
Tutor feedback	Substantive comments	.43	.06	.19
	Solve frequency	.33	<.01	.11
	Interaction episodes	.35	<.01	.12

were no direct relationships between the frequency of tutees' incorrect statements or the tutors' feedback and the observers' learning.[3]

Despite the previously described pattern of correlations, the possibility remains that it is the content of the conflicts per se that enabled the observers to learn; that is, it is the misconceptions expressed and the refutations that provided correct information that may have helped observers learn. Thus, a second way to test our hypothesis that it is not the content of the incorrect statements plus the refutation statement per se that enhances the learning of the dialogue-observers but rather that it is the conflicts themselves that encourage more active learning is to see whether a direct relationship exists between conflicts and observers' constructive behavior in another way. We compared the mean number of substantive comments, as originally coded in the Muldner et al. (2014) study, generated by the dialogue-observers (M = 8.39, SD = 4.84) for the 27 problem scenarios that contained conflict episodes with the mean number of substantive comments for the 43 scenarios that did not contain conflict episodes (M = 5.28, SD = 5.26) and found a significant difference, $F(1, 68) = 6.15$, $p = .02$ (see Figure 3). This confirms our interpretation that observers are more constructive (in generating substantive comments) during conflict episodes when errors are expressed.

Third, to test our interpretation about the role of conflict episodes in eliciting more active engagement from dialogue-observers even more directly, we compared the mean number of substantive comments observers made right before a tutee articulated an incorrect statement (2.90 substantive comments) versus right after the tutee articulated an incorrect statement (3.72), and the difference was

[3]Ideally a mediation model analysis would be used to further investigate whether observers' behaviors analyzed in these correlations influenced the effect of tutee moves on observers' learning. However, the sample size was not large enough for such a model.

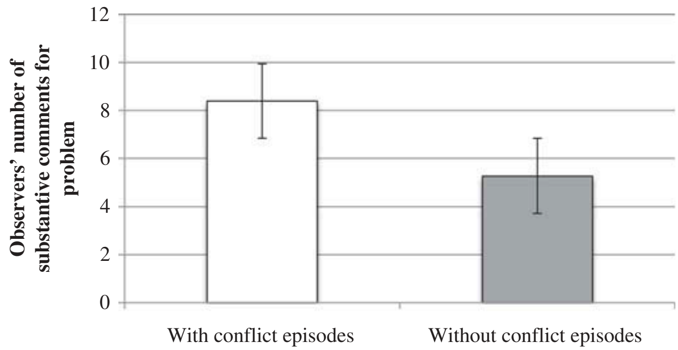

FIGURE 3 Number of substantive comments generated for problems with and without conflict episodes. Error bars are ±2 *SE*.

again significant, $\chi^2(1, N = 78) = 3.95, p < .05$. This shows that observers were more constructive after they saw conflicts.

In summary, we have provided some evidence for the novel hypothesis that seeing conflict episodes encourages observers to be more actively engaged by being more constructive and interactive, such as generating more substantive comments during conflict episodes and right after an incorrect statement is articulated by tutees.

DISCUSSION AND CONCLUSION

The goal of this article was to explain why tutorial dialogue-videos enhanced observing students' learning more so than lecture-style monologue-videos. We explored three sets of explanations by analyzing video materials and students' process data that were collected in a prior study (Muldner et al., 2014) that showed the results of the main finding that we are trying to explain: specifically, that tutorial dialogues are superior to lecture-style monologue-videos as instructional formats for student learning. The first set of hypotheses focused on the content of the videos in terms of the tutors' and tutees' moves, such as tutors' elaborative feedback and the deep questions they asked, the frequency of tutors' iconic gestures, the extent of concept coverage, the presence of misconceptions expressed by tutees, and so forth; none of these moves yielded any significant correlations with observing students' learning. Thus, it seems that these video content factors can be ruled out as possible explanations to account for the advantage of dialogue-videos for observers.

Instead, we postulated a second set of hypotheses relating to differences in the behaviors the observing students themselves displayed while watching dialogue-videos and monologue-videos. Indeed, we found significant differences in the constructive and interactive behaviors of the observing students when watching dialogue-videos compared to watching monologue-videos. These behavior differences were frequency of solving (vs. copying) problems, frequency of substantive comments generated, and frequency of coconstructive interactions. Thus, watching dialogue-videos prompted the observing students to be more constructive and interactive compared to watching monologue-videos.

We then postulated a novel third hypothesis for why dialogue-videos triggered, without explicit prompts, more constructive and interactive behaviors from the observing learners. We hypothesized that the dialogue-observers benefited from the presence of the tutees, and we showed that they did in fact pay more attention to what the tutees said than to what the tutors said. It is counter-intuitive that even though tutees articulated incorrect statements, whereas tutors presumably articulated only correct statements, dialogue-observers learned more even though they repeated and elaborated more often on what the tutees said compared to the monologue-observers, who necessarily repeated and elaborated on what the tutor said. Also, repeating is a kind of active engagement, whereas elaborating is a more advantageous constructive engagement, so the monologue-observers' greater repetitions of the tutors' correct information (compared to the dialogue-observers' repetitions) were not helpful to them. This result further confirms that tutees' statements are important learning resources for dialogue-observers.

In addition to the social presence provided by the tutees, two additional concrete explanations can be provided for the benefit of their presence. The first explanation is that tutees can serve as a model of learning. This is supported by our finding showing that the observing students seem to imitate tutees' constructive behaviors, such as generating substantive comments and asking questions, which are important constructive learning skills.

The second concrete advantage of the tutees' presence is that tutees tend to make errors and struggle, which is followed by tutor feedback. We call such cycles of errors followed by feedback *conflict episodes*. We had proposed that the benefit of seeing tutees struggle is more than knowing that the tutees' answers are incorrect, and therefore they should not rely on or copy the tutees' answers or explanations but should resort to their own effort to solve. Instead, we speculate that seeing conflict episodes might encourage and motivate students to try harder. Such motivation could be triggered by a sense of empathy or a fear of failing so that the observers want to try harder. There is some evidence in the literature suggesting the validity of this empathy interpretation. Chase, Chin, Oppezzo, and Schwartz (2009) asked eighth-grade students to instruct a character called Teachable Agent in a computer-based learning environment. The Teachable Agent reasoned based on how it was taught by the students. Students participated in one of two conditions: They were told that they were either teaching an agent or teaching for themselves. Students working with a teachable agent empathized with their agent in that they wanted to make sure their agent learned. They did this by putting in greater effort, such as in revising their concept maps, a constructive activity. That is, they treated their agent as their protégé. In contrast, teaching an avatar of oneself did not result in the same level of effort. We might make the analogy that the observing learners are in a comparable situation as students teaching a Teachable Agent, in that the observing students may empathize with the tutees and wish not to fail or err, as their tutees do.

We might also ask why observing students prefer to attend to the tutee more so than the tutor in the videos. This result has now been found in two separate sets of data. One explanation is that the tutee is a novice learner, much like the observing students. Being alike, they share a similar naïve understanding, or what we have previously called the *zone of representational match* between the tutees and the observing students (Chi, 2013), playing off the notion of the zone of proximal development. The zone of representational match simply means that because experts (e.g., the tutors/instructors) and novice learners (e.g., the tutees and the observers) have very different representations of a problem (Chi, Feltovich, & Glaser, 1981), the observing students' representation/understanding is more likely to align with the understanding of the tutee (because they are both novice learners) than with the tutor's representation. Thus, because there is a greater mismatch or misalignment with the tutor's representation, it is more difficult for observers to understand what the tutors ask, express, and explain compared to what the tutees ask, express, and explain. We found evidence of this lack of alignment in the context of a lack of convergence between the tutor's normative mental model and the tutee's naïve mental model or understanding in the context of tutoring (Chi, Siler, & Jeong, 2004). Thus, the supposition that conversations require an alignment of representations between the two speakers (Garrod & Pickering, 2004) may be true at a superficial level, but there is no evidence in our data of a deeper alignment or convergence toward a deep common representation between a tutor and a tutee.

Our interpretation of the ease of understanding an aligned novice representation is further reinforced by a finding in Schunk et al. (1987), who showed that observing students judged themselves to be similar in competence to a struggling model, which suggests that they may understand a tutee (a struggling model) better than a tutor (an expert/sailing model). In short, it seems that observing students may understand the tutees' contributions (such as comments and questions) better than the tutors' contributions, which thus explains why they pay more attention to the tutees than the tutors. There may be other reasons why dialogue-videos are easier to understand that are not considered in this article. For example, conversations are fragmentary and elliptical, dialogues contain more pauses, the information conveyed is less dense per minute, and so forth.

The findings from this work are enticingly promising because the observing students' constructive and interactive behaviors were naturally elicited. This suggests that the advantage of tutorial dialogues can be reused and scaled up as videos for online learning that can maximize students' learning. Currently online learning materials are mostly monologue-based with a talking head, but even if they include multimedia, such as demonstrations, visual displays, and so forth, such presentations nevertheless entail a single instructor explaining demonstrations and visual displays in a lecture style, much as our tutors/instructors did in our monologue-videos. We suggest instead that dialogue-based videos might enhance students' learning to a greater extent.

In Chi (2009), Chi and Wylie (2014), Fonseca and Chi (2011), and Menekse et al. (2013), we proposed and demonstrated that ICAP can serve as a tool for resolving discrepant findings in the literature, for dictating what an appropriate control condition is, and for coding transcripts and other student products. The analyses presented here further illustrate how ICAP can guide the design of a new learning environment as well as provide evidence in support of ICAP's assumption that higher modes of engagement, the constructive and interactive modes, are responsible for deeper learning, as reflected in the deeper transfer questions we used in our assessment of learning. Most important, ICAP, as a theory of learning, also has the potential to be translated into a theory of instruction. We have just completed a 4.5-year project in which we used ICAP to provide prescriptions to teachers for how to modify and improve their worksheet activities in order to engage students at higher modes. Our attempt at translating a theory of learning into a theory of instruction not only is promising but suggests that minor tweaks in the design of teachers' worksheet activities might suffice to elicit higher modes of engagement from students, potentially leading to deeper learning. Combining the findings from that project with those reported in the current analyses could yield powerful guidelines for educators and instructional designers for generating robust online learning materials generalizable across domains.

ACKNOWLEDGMENTS

We appreciate the comments and editing provided by Joshua Adams, Christiana Bruchok, Emily Bogusch, Na Li, and Dongchen Xu. Special thanks go to Kasia Muldner for her work on the prior 2014 study.

FUNDING

We are grateful for funding for this project provided by the Spencer Foundation (Grant No. 200800196) as well as the National Science Foundation (Grant No. 1504893).

REFERENCES

Anderson, J. R., Corbett, A. T., Koedinger, K. R., & Pelletier, R. (1995). Cognitive tutors: Lessons learned. *Journal of the Learning Sciences*, *4*, 167–207.

Berkowitz, M. W., & Gibbs, J. C. (1982). Measuring the developmental features of moral discussion. *Merrill-Palmer Quarterly*, *29*(4), 399–410.

Biocca, F., Harms, C., & Burgoon, J. K. (2003). Toward a more robust theory and measure of social presence: Review and suggested criteria. *Presence: Teleoperators and Virtual Environments*, *12*(5), 456–480. doi:10.1162/105474603322761270

Bloom, B. S. (1984). The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational Researcher*, *13*(6), 4–16. doi:10.3102/0013189X013006004

Chan, C., Burtis, J., & Bereiter, C. (1997). Knowledge building as a mediator of conflict in conceptual change. *Cognition and Instruction*, *15*(1), 1–40. doi:10.1207/s1532690xci1501_1

Chase, C. C., Chin, D. B., Oppezzo, M. A., & Schwartz, D. L. (2009). Teachable agents and the protégé effect: Increasing the effort towards learning. *Journal of Science Education and Technology*, *18*(4), 334–352. doi:10.1007/s10956-009-9180-4

Chi, M. T. H. (1997). Quantifying qualitative analyses of verbal data: A practical guide. *Journal of the Learning Sciences*, *6*, 271–315.

Chi, M. T. H. (2009). Active-constructive-interactive: A conceptual framework for differentiating learning activities. *Topics in Cognitive Science*, *1*(1), 73–105. doi:10.1111/j.1756-8765.2008.01005.x

Chi, M. T. H. (2013). Learning from observing experts. In J. J. Staszewski (Ed.), *Expertise and skill acquisition: The impact of William G. Chase* (pp. 1–28). New York, NY: Psychology Press.

Chi, M. T. H., de Leeuw, N., Chiu, M. H., & LaVancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science*, *18*(3), 439–477.

Chi, M. T. H., Feltovich, P., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science*, *5*(2), 121–152. doi:10.1207/s15516709cog0502_2

Chi, M. T. H., Roscoe, R., Slotta, J., Roy, M., & Chase, M. (2012). Misconceived causal explanations for emergent processes. *Cognitive Science*, *36*(1), 1–61. doi:10.1111/j.1551-6709.2011.01207.x

Chi, M. T. H., Roy, M., & Hausmann, R. G. H. (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. (2004). Can tutors monitor students' understanding accurately? *Cognition and Instruction*, *22*(3), 363–387. doi:10.1207/s1532690xci2203_4

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

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

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2013). *Applied multiple regression/correlation analysis for the behavioral sciences*. New York, NY: Routledge.

Cohen, P. A., Kulik, J. A., & Kulik, C. L. (1982). Educational outcomes of tutoring: A meta-analysis of findings. *American Educational Research Journal*, *19*(2), 237–248. doi:10.3102/00028312019002237

Craig, S., Chi, M. T. H., & VanLehn, K. (2009). Improving classroom learning by collaboratively observing human tutoring videos while problem solving. *Journal of Educational Psychology*, *101*(4), 779–789. doi:10.1037/a0016601

Craig, S., Driscoll, D., & Gholson, B. (2004). Constructing knowledge from dialog in an intelligent tutoring system: Interactive learning, vicarious learning, and pedagogical agents. *Journal of Educational Multimedia and Hypermedia*, *13*(2), 163–183.

Craig, S., Gholson, B., Ventura, M., Graesser, A. C., & Tutoring Research Group. (2000). Overhearing dialogues and monologues in virtual tutoring sessions: Effects on questioning and vicarious learning. *International Journal of Artificial Intelligence in Education*, *11*(1, 242–253.

Craig, S., Sullins, J., Witherspoon, A., & Gloslon, B. (2006). The deep-level-reasoning-question effect: The role of dialogue and deep-level-reasoning questions during vicarious learning. *Cognition and Instruction*, *24*(4), 565–591. doi:10.1207/s1532690xci2404_4

Daradoumis, T., Bassi, R., Xhafa, F., & Cabelle, S. (2013). A review on massive e-learning (MOOC) design, delivery, and assessment. In F. Xhafa, L. Barolli, D. Nace, S. Vinticinque, & A. Bui (Eds.), *2013 Eighth International Conference on P2P, parallel, grid, cloud and Internet computing (3PGCIC)* (pp. 208–213). Compiegne, France: IEEE.

Driscoll, D. M., Craig, S., Gholson, B., Hu, X., & Graesser, A. (2003). Vicarious learning: Effects of overhearing dialog and monologue-like discourse in a virtual tutoring session. *Journal of Educational Computing Research*, *29*(4), 431–450. doi:10.2190/Q8CM-FH7L-6HJU-DT9W

Fonseca, B., & Chi, M. T. H. (2011). The self-explanation effect: A constructive learning activity. In R. E. Mayer & P. A. Alexander (Eds.), *The handbook of research on learning and instruction* (pp. 296–321). New York, NY: Routledge.

Garrod, S., & Pickering, M. J. (2004). Why is conversation so easy? *Trends in Cognitive Sciences*, *8*(1), 8–11. doi:10.1016/j.tics.2003.10.016

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/(ISSN)1099-0720

Hew, K. F., & Cheung, W. S. (2014). Students' and instructors' use of massive open online courses (MOOCs): Motivations and challenges. *Educational Research Review*, *12*, 45–58. doi:10.1016/j.edurev.2014.05.001

McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. Chicago, IL: University of Chicago Press.

Menekse, M., Stump, G., Krause, S., & Chi, M. T. H. (2013). Differentiated overt learning activities for effective instruction in engineering classrooms. *Journal of Engineering Education*, *102*(3), 346–374. doi:10.1002/jee.v102.3

Monaghan, P., & Stenning, K. (1998). *Learning to solve syllogisms by watching others' learning* (Research Paper No. HCRC/RP-98). Edinburgh, UK: Human Communication Research Centre, University of Edinburgh.

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

Muller, D. A., Bewes, J., Sharma, M. D., & Reimann, P. (2008). Saying the wrong thing: Improving learning with multimedia by including misconceptions. *Journal of Computer Assisted Learning*, *24*(2), 144–155. doi:10.1111/j.1365-2729.2007.00248.x

Muller, D. A., Sharma, M. D., Eklund, J., & Reimann, P. (2007). Conceptual change through vicarious learning in an authentic physics setting. *Instructional Science*, *35*(6), 519–533. doi:10.1007/s11251-007-9017-6

Narciss, S. (2007). Feedback strategies for interactive learning tasks. In J. M. Spector, M. D. Merrill, J. van Merrienboer, & M. P. Driscoll (Eds.), *Handbook of research on educational communications and technology* (3rd ed., pp. 125–144). Mahwah, NJ: Erlbaum.

Okita, S. Y., Bailenson, J., & Schwartz, D. L. (2007). The mere belief of social interaction improves learning. In D. S. McNamara & J. G. Trafton (Eds.), *The proceedings of the 29th Annual Meeting of the Cognitive Science Society* (pp. 1355–1360). Mahwah, NJ: Erlbaum.

Rummel, N., & Spada, H. (2005). Learning to collaborate: An instructional approach to promoting collaborative problem solving in computer-mediated settings. *Journal of the Learning Sciences*, *14*, 201–241. doi:10.1207/s15327809jls1402_2

Schunk, D. H., Hanson, R. A., & Cox, P. D. (1987). Peer-model attributes and children's achievement behaviors. *Journal of Educational Psychology*, *79*(1), 54–61. doi:10.1037/0022-0663.79.1.54

Stenning, K., McKendree, J., Lee, J., Cox, R., Dineen, F., & Mayes, T. (1999). Vicarious learning from educational dialogue. In C. M. Hoadley & J. Roschelle (Eds.), *Proceedings of the 1999 Conference on Computer-Support for Collaborative Learning* (pp. 341–347). Palo Alto, CA: Stanford University.

Tversky, B., Jamalian, A., Segal, A., Giardino, V., & Kang, S. (2014). *Congruent gestures can promote thought*. Toronto, ON: Association for Computing Machinery Special Interest Group on Computer-Human Interaction.

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., Lynch, C., Schulze, K., Shapiro, J. A., Shelby, R., Taylor, L., … Wintersgill, M. (2005). The Andes physics tutoring system: Lessons learned. *International Journal of Artificial Intelligence and Education*, *15*(3), 1–47.

APPENDIX A: RUBRIC OF 20 CONCEPTS RELEVANT TO THE CONCEPT OF DIFFUSION AND EXAMPLES OF HOW THEY WERE STATED IN VIDEOS BY THE INSTRUCTORS/TUTORS

1. Micro Continuous Movement

Molecules are always moving. Molecules have continuous movement. They never stop moving. Also may include molecules do not lose kinetic energy upon an elastic collision. They do not speed up or slow down over time unless there is a change in temperature. Molecules do not stop moving at equilibrium.

2. Micro Universal Behavior

All molecules behave the same way. Molecules behave the same regardless of if they are dye, water, CO_2, O_2, or any other type.

3. Micro Movement

Molecules move around, bounce off each other, collide, move quickly.

4. Micro Dissolve

Molecules do not dissolve. They do not decompose, separate into smaller parts, become one with other molecules, or create a new substance. Also includes how flow patterns are not from molecules dissolving.

5. Micro Unpredictability

Molecules move randomly. Their direction cannot be predicted at any point in time. We cannot make a guess as to where molecules will move.

6. Micro Inanimate

Molecules are inanimate. They do not have feelings, wants, or needs. They do not need or want to reach equilibrium or move with a flow.

7. Macro Flow

Flow of molecules tends to be from high concentrations to low concentrations. This concept is not connected to unpredictable movement or micro behavior.

8. Macro Equilibrium

Equilibrium is when molecules are evenly spread out, concentrations are about equal throughout a solution, it is homogeneous, it is all one color.

9. Macro Diffusion

Over time a mixed solution will reach equilibrium. Only discusses time and homogeneity. Does not connect with random movement or micro behavior.

10. Macro Predictability

Flow is a predictable behavior.

11. Macro Visibility

Flow is visible to the naked eye. Flow is what we see. Also the macro pattern is visible.

12. Macro/Micro Independence

Molecule movement is independent of flow patterns. Molecules may move backwards against the flow. What we see is not necessarily how molecules behave.

13. Macro/Micro Flow

Connects how random movement of molecules is connected with flow patterns. If there is a lower concentration of a substance in an area, and molecules move randomly and continuously, there is a greater chance the molecules moving any which way will end up there. This results in a visible flow.

14. Macro/Micro Diffusion

Connects random movement of molecules with how they spread out. Random and continuous movement of molecules causes them to spread out and reach equilibrium. Makes no mention of greater chances to flow to less concentrated areas or flow patterns.

15. Micro Concentration

Molecule behavior is independent of concentrations they are in. Molecules do not behave any differently at equilibrium or difference concentrations.

16. Concentration Definition

The definition of concentration is a measure of the amount of a certain substance in a solution relative to the total amount. Concentration is not dependent solely on the amount.

17. Compute Concentration

The process of finding concentration. Showing or telling how to take a fraction for concentration, counting up molecules, using the total amount of all molecules, writing it as a percent.

18. Macro Stable Equilibrium

A solution at equilibrium stays at equilibrium if left alone. There is no flow at equilibrium.

19. Gravity

Molecules are not affected by gravity (in this material). Gravity does not pull molecules downwards or affect how they spread out.

20. Macro/Micro Equilibrium Balance

Equilibrium is not a perfect balance. Since molecules are constantly moving, at times there will be very slight unbalances.

APPENDIX B: EXAMPLES OF INCORRECT TUTEE STATEMENTS, THE CONCEPT TO WHICH THEY REFER, ALONG WITH THE SUBSEQUENT DISCOURSE

Example 1

S: <u>They're [the molecules] trying to reach equilibrium</u>. [Concept 6: Micro Inanimate]

T: okay kind of—the key word here is this word need …

S: oh okay
T: so dye molecules don't need to do anything.
S: umhum
T: Again they don't really care if the state is in equilibrium they are just gonna continue their random motion.
S: okay

Example 2

S: You can't really tell, can you? [when in fact you can] [Concept 5: Micro Unpredictability]
T: not quite
S: (giggle) ummm
T: So let me let me give you let's go back to the other question (brings pre slide) to help us understand. So over here we said this would flow to the right and why would it flow to the right.

Example 3

S: Yeah, I think they're going faster. [Concept 1: Micro Continuous Movement]
T: … we reset to see if there is a speed difference … so just wait for a second … so you said that it is getting faster so like throughout the time.
S: Yeah
T: The dye and water molecules initially they are slower.
S: Yeah
T: So let's see, do you think they are getting speed up?
S: Yeah it looks like it sort of
T: Yeah (dubious) maybe it is … just an illusion of
S: yeah they are all going together
T: Actually they are in the same speed form start to end and there is no end by the way they keep moving
S: okay

Example 4

S: I think it would be random … [Concept 10: Macro Predictability]
T: It would be random
S: Yeah I'm mean I'm not really sure but as far as which way I'm not sure which way it would go, but I think it would still you know well I guess no actually the flow it would be the same, because it was in one standard direction of … they are staying outside the cell they are not going inside.
T: So we have highly concentrated oxygen

S: yeah
T: Outside
S: Outside
T: But, inside of the cell we have not concentrated oxygen, okay

Note. Underlining indicates incorrect statements. T = tutor; S = tutee.

APPENDIX C: EXAMPLES OF INTERACTION AND NONINTERACTION EPISODES

	Line		
Interaction Episode 1 (Molecules Lack Intent)	1	**F**: They don't have a specific goal. They're just all over the place	*Turn 1*
	2	**M**: They don't have a flow	*Turn 2*
	3	**F**: and they don't stop moving when they reach equilibrium. // So it's not their goal or whatever cause they have a mind. (Laughs)	
	4		
	5	**M**: That's funny!	
Interaction Episode 2 (Molecules Move Randomly)	1	**F**: There's more room to move (pause video)	*Turn 1*
	2	**M**: They're not trying to find room. There's just…they just have to move. So they eventually get.	*Turn 2*
	3		
	4	**F**: Wait, there's more, there is more room to move	*Turn 3*
	5	**M**: I didn't say that [there is more room to move]	*Turn 4*
	6	**F**: I know. I did because she said that [there is more room to move]	
	7	**M**: Can you rewind that?	*Turn 5*
	8	**F**: Yeah.	
	9	**M**: So just be like there's not a specific destination…like they just go there because like they bounce in there.	
	10		*Turn 6*
	11	**F**: and there's a random…	
	12	**M**: Yeah	

(*Continued*)

(*Continued*)

Interaction Episode 3 (Molecules do not Dissolve)	1	**M**: They're just coexisting (pause video)	*Turn 1*
	2	**F**: They didn't say an answer yet.	
	3	**M**: They're not co…they're not dissolving into each other. They're just com…mixing with each other	
	4		
	5	**F**: Coexisting and mixing?	
Noninteraction Episode 4* (Computing Concentration)	1	**N**: ok so there are similar (writing) // should be… mutter (6 over 17) / they are the same (mutters as writes) right?	
	2		
	3	**T**: ah ha	
	4	**N**: [v] and oh I should be c1	
NoninteractionEpisode 5* (Equilibrium)	1	**H**: (reads first statement) because you can have more water than dye //so it can never it might never reach equilibrium. Like how in salt water you can add salt // but then it still wont see water so it's never even	
	2		
	3		
	4	**Y**: Oh, ok so…	
	5	**H**: So it is even with dye solution might not - like the dye could just6 sink to the bottom (continues writing)	
	6		
Noninteraction Episode 6* (Observable Pattern of Diffusion)	1	**K**: okay, let's do this. I think, I was gonna say the same thing she was saying like the first second like the ink is like barely getting into it	
	2		
	3	**Z**: uhum	
	4	**K**: but it's at still like gathered around (can't hear) – (writing)	
	5	**Z**: yeah	
	6	**K**: Something like that	
	7	**Z**: Mmhm	

***Non-interacting episodes (3-6) contain no turns**

Note. Each turn is a substantive turn. F, M, N, T, H, Y, K, and Z denote various speakers.