HAL
open science

An Approach to Analyzing the Role and Structure of Reflective Dialogue

Sandra Katz, Gabriel O'Donnell, Heather Kay

To cite this version:

Sandra Katz, Gabriel O'Donnell, Heather Kay. An Approach to Analyzing the Role and Structure of Reflective Dialogue. International Journal of Artificial Intelligence in Education, 2000, 11, pp.320-343. hal-00197333

HAL Id: hal-00197333

https://telearn.hal.science/hal-00197333v1

Submitted on 14 Dec 2007

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

International Journal of Artificial Intelligence in Education (2000) 11, 320-343

An Approach to Analyzing the Role and Structure of Reflective Dialogue

Sandra Katz, Gabriel O'Donnell, Heather Kay *Learning Research And Development Center, University Of Pittsburgh, Pittsburgh, Pa 15260 USA*
Email: katz+@pitt.edu

Abstract. Several intelligent tutoring systems contain modules that support reflection after practice exercises. However, there is very little research on reflective dialogue and its relation to dialogue during problem solving that can guide this effort. The approach to analyzing educational dialogue described in this paper was developed to address this problem. Using statistical and machine learning methods, we compared the frequency and structure of speech acts by avionics students and experts during practice exercises with speech acts during a post-practice reflective discussion (PPR). There were several differences. For example, the percentage of student requests for explanations about how the electronic system works and of self-explanations was higher during PPR than during problem solving. We also investigated the relationship between when a misconception was addressed—during problem solving, during post-practice reflection, or distributed between the two phases—and whether a misconception was eliminated. Misconceptions were more likely to be resolved if they were addressed during problem solving and PPR than during problem solving alone. These results suggest that PPR can play an important role in instructing students in the conceptual knowledge underlying strategic tasks, and hence support the inclusion of reflective activities in intelligent tutoring systems.

INTRODUCTION

In traditional apprenticeships, the "master" typically guides the apprentice while the apprentice carries out a complex task. The master models the correct way of doing things, "scaffolds" difficult parts of the task by taking over when necessary, and fades support as the learner gains competence (Collins & Brown, 1989; Collins, Brown, & Newman, 1989). But a significant part of apprenticeship training takes place *after* a task, during a reflective conversation referred to by many names such as the "debrief" or "postmortem." We will refer to this conversation as *post-practice reflection* (PPR), to highlight its temporal and instructional aspects. During PPR, the master typically comments on critical aspects of the learner's performance, as when a football coach replays a videotape of a rookie's game and pauses to give feedback on important plays.

There is strong evidence that reflection *during* instructional activities is important (e.g., Bielaczyc, Pirolli, & Brown, 1993; Chi & VanLehn, 1991; VanLehn, Jones, & Chi, 1992). However, as other researchers have noted (e.g., Lederman, 1992), there is very little research on *post*-practice dialogue and its role in instruction. This is surprising, given that apprenticeship training and debrief are as old as training itself in some domains, such as the military and health professions. In recent years, several developers of intelligent tutoring systems (ITS) have developed modules to support reflection (e.g., Bull & Broady, 1997; Katz, Lesgold, Hughes, Peters, Gordin & Greenberg, 1998; Pioch, Roberts, & Zeltzer, 1997; Roberts, 1993). Hence, research investigating what happens during reflective dialogues, whether these dialogues support learning, and (if so) how this instructional effect occurs, is needed to guide the design of apprenticeship-style training programs—computer-based or otherwise.

Prior research on post-practice reflection has described its critical components and phases, demonstrated its positive impact on team training, and begun to reveal how the structure of reflective explanations support learning. Lederman (1992) described the main elements of the debriefing process as it takes place in instructional simulations and games: the guide, participants (e.g., students), an experience, and discussion of the impact of that experience.

Lederman also derived a model of the key phases of debrief in these settings. Smith-Jentsch, Payne, and Johnston (1996) demonstrated that team leaders can be trained to conduct effective pre-briefs and debriefs of tasks involving teamwork, and found that teams so trained demonstrated better teamwork and outperformed a control group. Moore (1996) analyzed the structure of human experts' reflective explanations in avionics, in order to develop a discourse planner for reflective dialogue. Moore, Lemaire, and Rosenblum (1996) specified the ways in which students and tutors refer to prior explanations during reflective dialogues. Rosé (1997) identified common student moves during reflective dialogues, such as explanations of their reasoning and "why" questions. Di Eugenio, Moore, and Paolucci (1997) used machine learning methods to specify characteristics of reflective explanations that predict what types of discourse markers tutors use to make an explanation coherent and understandable. Katz, Aronis, and Creitz (1999) used machine learning to distinguish problem-solving dialogue from reflective dialogue with respect to speech act frequency and structure, and to interpret the instructional role of particular speech acts in particular positions.

This article summarizes our comparative analysis of problem-solving and reflective dialogue and provides experimental support for including PPR in apprenticeship-style training programs. The main instructional issue to be addressed is: *what role, if any, does reflective dialogue play in eliminating misconceptions?* We first describe the approach to educational dialogue analysis that we developed to compare the two phases of instruction, to study their relationship, and to identify the role of PPR in learning. Our analyses are based on a corpus of tutorial dialogues that took place between avionics experts and collaborating pairs of students (dyads) who worked on practice exercises in **Sherlock 2**, an ITS for electronic fault diagnosis (e.g., Katz, Lesgold, Eggan, & Gordin, 1993; Katz et al., 1998; Lesgold, Eggan, Katz, & Rao, 1992).

APPROACH TO EDUCATIONAL DIALOGUE ANALYSIS

Dialogue corpus

The sample of problem-solving and reflective dialogues was collected during a formative evaluation of **Sherlock 2.** Our goal was to determine how the automated coach would need to be modified to support collaborating students (Katz, 1995; Katz & O'Donnell, 1999). We observed eight pairs of avionics students (dyads). Each dyad was matched with one of eight experienced avionics technicians from a local Air National Guard unit. Students took turns "coaching" their peer. In apprenticeship fashion, the mentors' roles were to guide the student coach when he could not help his peer and to comment on students' performance after they solved a problem.[i]

Interaction during problem solving was typed into a "chat" window and was spoken during PPR. The difference in communication mode occurred because we included the debrief activity as an afterthought, to solve a problem with the experimental setting. Since the peer coach was often able to advise his partner, the mentor could be inactive for long periods. Preparing for the debrief sustained the mentors' attention. Hence the data provided us with many samples of reflective dialogue (approximately thirty-five sessions, about twenty minutes each), although studying reflective dialogue and its relation to coaching during problem solving was not our original intent.

Overview of the approach

Like several approaches to dialogue analysis (e.g., Pilkington, 1997), our approach is hierarchical. (See Figure 1 and, for more detail, Appendix 1. Figure 1 is based on the sample dialogue shown in Table 2. Due to space constraints, several nodes in Figure 1 are abbreviated, and some coding fields were eliminated.) We describe speaker intent using a taxonomy of speech acts tailored to instructional dialogue, as shown in Table 1. Some of the speech acts are

"atomic," because they are not subdivided further—that is, *Advise*, *Disclose*, *Question*, *Appraise*, *Direct*, *Make-Sense*, *Instruct*, and their members. Other speech acts—that is, *Explain*, *Justify*, and their members—are "complex" because they consist of atomic speech acts and, recursively, other complex speech acts. We specify speech acts further by describing their relationships with other speech acts—for example, *Knowledge Explanation* of an *Advise* act, which we write as "*Knowledge Explanation* of advice."

We parse each speaker's dialogue turn (*contribution*) into its constituent speech acts (*contribution segments*). The intermediate nodes of the parse tree consist of complex speech acts, while the leaves are atomic speech acts. The atomic components of a complex speech act are called *contribution subsegments*. To represent dialogue interactions, we group speech acts into *exchanges*. An exchange consists of one speaker's initiating speech acts about an instructional theme—that is, a concept or strategic rule—followed by another speaker's response. A series of exchanges that address the same theme forms an *instructional dialogue* (ID). At the highest level, shown as a blank node at far left of Figure 1, instructional dialogues belong to *conversations.* A conversation consists of one or more ID's and a closing speech act or problem-solving action.

Although the units of analysis at higher levels inherit information recorded at lower levels, we can treat the levels independently and focus on specific questions at each level. We illustrate this point below for the three levels of the hierarchy that we have focused on to date: contribution segments, exchanges, and instructional dialogues. We will refer to the conversations in Figure 1 and Appendix 1 for examples. (For brevity, the lowest level of analysis, contribution subsegment, was eliminated from Figure 1 and Appendix 1, except for the sample subsegments shown in Appendix 1, lines 145-159.)

Table 1: Speech Acts During Learning Interactions

Advise
Clarify
Disclose
Question
  Request
  Disjunctive-request
  Diagnostic-query
Make-sense
  Interpret
  Convert
  Predict
  Hypothesize
Instruct
  Inform
  Define
  Identify
  Compare
Hint
Appraise
  Confirm
  Refute
  Correct
Explain
  Step explanation
  Strategic goal explanation
  Knowledge explanation
  Causal process description
Justify
  Step justification
  Strategic goal justification
  Knowledge justification
Other

Contribution segments

Speech acts represent discourse pragmatics (Searle, 1969): what a speaker is trying to do with an utterance. Each speech act is defined in terms of a dialogue goal—e.g., *Advise*, in order to direct action; *Explain*, in order to enable understanding. The speech act categories in Table 1 map onto the domain-neutral descriptors of speaker intent in other dialogue-coding systems—e.g., *Instruct* maps to *Enable* in Rhetorical Structure Theory (Mann & Thompson, 1988). Similarly, our taxonomy of topics map onto general semantic categories—that is, actions, states, and events—and are specified according to the task domain—e.g., *measurement* action. Our approach thus supports the modeling of educational discourse from general and domain-specific perspectives.

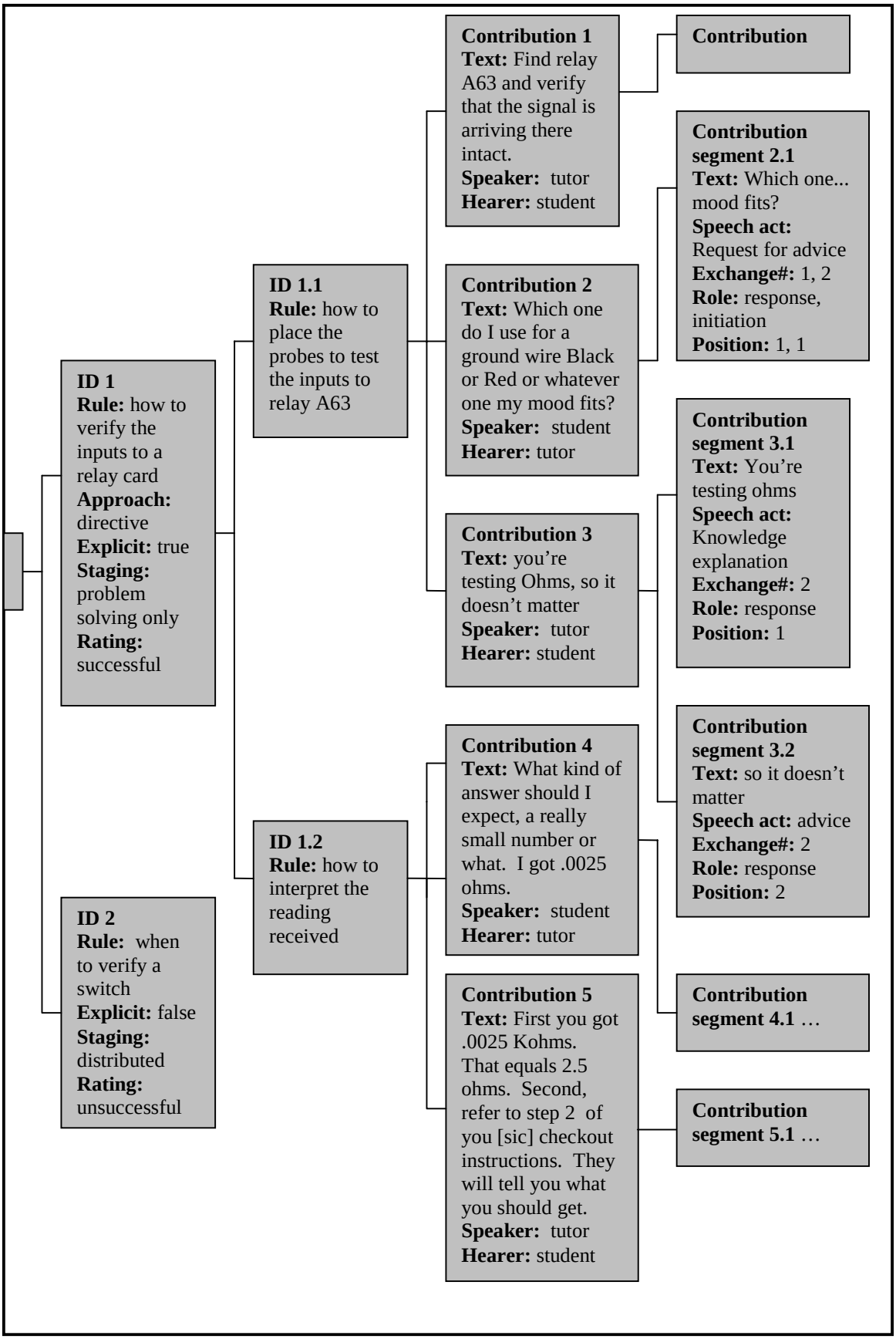

Figure 1. A Hierarchical Approach to Dialogue Analysis

The taxonomy of speech acts shown in Table 1 was influenced by Stiles and Putnam's (1992) classification of speech acts for analyzing doctor-patient interactions. We make finer distinctions within some categories, to gain more descriptive power. For example, we adopted Chandrasekaran, Tanner, and Josephson's (1989) classification of explanations in diagnostic domains: explanations about how a system works (*Knowledge Explanations*), explanations about system status according to the results of diagnostic actions (*Step Explanations*), descriptions of causal mechanisms (*Causal Process Descriptions*), and explanations about how and why to achieve particular problem-solving goals (*Strategic Explanations*).

We assign a speech act label to each contribution segment, as illustrated in Figure 1 and Appendix 1 (*Speech Act* field). In addition, we record several linguistic, semantic, and extra-linguistic attributes. Person is a linguistic feature, and is coded as a numeric flag: 1 = first person, 2 = second person, and 3 = another dialogue participant such as the experimenter or Sherlock, the automated coach. For example, "*Strategic Justification* of request-1 for advice-2" (Appendix 1, lines 28-30) means A strategic justification of *my* request for advice from *you*." (The first speech act in the contribution segment field is assumed to be by the speaker, so first person is not marked.).

With respect to semantics, we specify the strategic rule, subrule, and/or piece of conceptual knowledge addressed by the tutor's or peer coach's speech acts (*Rule*, *Subrule*, and *Conceptual knowledge* fields, respectively). Rules and subrules correspond to goals and subgoals, respectively, in the task hierarchy, as illustrated in Table 2. This goal-based information allows us to encode the instructional import of speech acts—for example, what strategic rule the tutor is *Advis*ing the student about. In the *Predicate* field, we represent the specific claim expressed about a domain referent—e.g., *Verify-component (TPS Relay Card)*, Appendix 1 lines 61-62, represents the tutor=s correction to verify the Test Point Select Relay Card, in contribution segment 2.1. Finally, we record several extralinguistic features such as who the speaker is (*Speaker* field) and who the hearer is (*Hearer* field). This information is attached at the contribution level, since it applies to all of the contribution's "child" segments. We also mark whether the proposition stated in a speech act is true or false; correct, partially correct, or incorrect (*Rating* field). This rating represents the accuracy of the tutor's instruction, or of a student=s claim.

At the contribution segment level, we have investigated pragmatic and structural differences between problem-solving and reflective dialogue, as measured by the relative frequency of particular speech acts. The results of this analysis are summarized in the section entitled, "Applying the Approach."

Exchanges

An exchange consists of one speaker's initiation followed by another speaker's response. Often more than one theme (strategic rule or associated concept) is addressed within a speaker turn (contribution), so a contribution can be a member of more than one exchange. For this reason, we attach exchange-level information to a contribution's segments, as shown in Figure 1 and Appendix 1 (e.g., lines 21-23). We record three types of information. In *Exchange#*, we specify the exchange's order within the session. In the *Role* field, we specify whether the speech act belongs to the initiation or the response. Many responses are, in turn, re-initiating, so they receive both role codes (e.g., Appendix 1, lines 64-66). Finally, in the *Position* field, we record the contribution segment=s order within the initiation or response.

The exchange level supports investigation of numerous questions about tutor-student interaction, such as: who is primarily responsible for initiating the instructional dialogues—the tutor or the student? Who keeps the dialogue moving forward, as measured by the proportion of reinitiating moves by the student and tutor? Is there any correlation between these features of the dialogue and learning?

Missing from our scheme is a way of describing an exchange unit as a whole, rather than in terms of its members. One of our goals in developing this approach is to identify *exchange types*, through methods similar to those described in Stiles and Putnam (1992), such as cluster analysis. This analysis may refine Sinclair and Coulthard's (1992) classification of exchange

types in educational dialogue—e.g., *Question/Answer*, *Statement/Counter-statement*. The resulting taxonomy of exchanges will support formal descriptions of instructional dialogues (ID's) and of how ID structure correlates with learning.

Table 2: Illustration of the hierarchical knowledge base and its reflection in dialogue

Segment of the Strategic Rule Hierarchy	**Instructional Dialogue 1: Advice to apply Rule 25.1**
Rule25: If the aircraft unit's outputs are good and the connecting wires between the unit and the testing station are good, *verify the inputs to the component containing the relay activated by the failed checkout procedure* [a relay card]. **Rule25.1** To verify the inputs to a relay card, *identify the input signals* in the schematics, *determine how to set the meter*, *determine how to place the probes*, and *interpret the status of the readings received*. **Rule25.1.1** To identify the inputs to a relay card, use the technical orders to find the "test point" pins on the Unit Under Test, and trace forward from these pins. **Rule25.1.2** To determine how to set the meter when testing the inputs to a relay card... **Rule25.1.3** To determine how to place the probes when testing the inputs to a relay card ... **Rule25.1.4** To interpret the status of a reading on the inputs to a relay card...	**Instructional Dialogue 1.1: How to determine how to place the probes to test the inputs to relay A63** (Rule 25.1.3) Tutor: Find relay A63 (because that's what you tested) and verify that the signal is arriving there intact. Student: Which one do I use for a ground wire Black or Red or whatever one my mood fits? Tutor: you're testing OHMS, so it doesn't matter {student takes measurement and gets a .0025 Kohms reading} **Instructional Dialogue 1.2: How to interpret the reading received** (Rule 25.1.4) Student: What kind of answer should I expect, a really small number or what? I got .0025 Ohms. Tutor: first, you got .0025 Kohms. That equals 2.5 ohms. Second, refer to step 2 of you (sic) checkout instructions. They will tell you what you should get.

Instructional Dialogues

An instructional dialogue (ID) consists of a series of exchanges, unified by the domain principles, rules, and concepts they address. Since avionics centers around strategic knowledge, the ID's in our corpus are described in terms of the main strategic rule they address. Strategic knowledge is decision-making knowledge. According to Gott (1990), it is knowledge about "how to decide what to do, and when."

As shown in Figure 1 and Appendix 1, a conversation may contain several ID's, each one addressing a different focal rule. Through embedded discussions of subrules, ID structure often mirrors the hierarchical structure of the knowledge base. This is illustrated in Table 2 and Appendix 1 ID's 2, 2.1, and 2.2. The exchanges that make up an ID are not necessarily contiguous and the same focal rule may be addressed across several ID's in a given session. We refer to a set of instructional dialogues that address the same rule as an *ID Series*. (See Table 3 for an example of an ID Series that spans problem solving and PPR. For brevity, we do not show lower levels of discourse analysis.) Furthermore, ID's may overlap, as when one ID ends in an exchange that reinitiates a new exchange and new ID. ID's 1 and 2 in Appendix 1 illustrate overlapping ID's.

At the ID level, we record features of the interaction that can not be tied to a particular exchange or to any of its constituent speech acts, such as:

1. *Explicitness:* Did the tutor state the rule explicitly, or did the student have to infer or "construct" the rule from the interaction? For example, if the goal of a strategic rule (the "right-hand side") was stated during an *Advise* act, were the appropriate conditions (the "left-hand side") also stated?
2. *Approach:* Was instruction in the rule directive, or did the tutor guide the student using a series of questions and prompts?
3. *Staging:* Was the rule addressed only during problem solving, only during PPR, introduced during problem solving and restated during PPR (with no new information provided), or introduced during problem solving and elaborated upon during PPR?

In the next section, we discuss a study in which we investigated the relation between staging of explanations across problem solving and PPR, and learning.

Table 3. A Sample ID Series Addressing Rule 51 (Appendix 2)

References are to the circuitry shown in Figure 2.

ID During Problem Solving	ID During Post-practice Reflection
Solver: The output of the card had a bad reading, but [the relay] was getting the proper voltage across. **Student coach:** Coach, what do you think? **Mentor:** *A15's circuit path flows thru (sic) all of the relays after B24.* **Student coach:** Try relay 23 pin 36 and ground.	**Mentor:** You checked this relay right here [pointing to relay B24 on card A1A3A15], which was good. Then you measured here [pointing to pins 11 and 12], and it was bad. **Solver:** Right **Mentor:** And it was good here [input pins 57 and 58]. Somehow, between here [inputs] and here [outputs] you were getting an open. And, you did check your signal which was coming from your A10 card to B24, which was good…The one thing you overlooked is that there are one, two, three four, five other relays that are getting power from your A10 card…If ground was to switch, and kick another rely on, then that would mess up your measurement path prematurely. So you replaced your A15 prematurely… **Solver:** OK

APPLYING THE APPROACH

In this section, we describe our first attempts to use the approach to address issues in educational dialogue analysis. In particular, we compared reflective dialogue with dialogue during problem solving, and investigated whether reflective dialogue had a positive effect on learning in the instructional setting described above. These studies demonstrate the potential of the approach and reveal some of its limitations. They also suggest that the PPR discussions enhanced the avionics students' acquisition of strategic and conceptual knowledge.

How do instructional interactions during problem solving compare with interactions during post-practice reflection?

As we discussed previously, the contribution segment level allowed us to identify pragmatic differences between problem solving and reflective dialogue. Two raters coded the data. Rater agreement on speech act codes was difficult to achieve and inconsistent; kappa scores ranged from .55 to .89. Differences were discussed and re-coded. By comparing the percentage of speaker turns that contained at least one instance of a particular speech act, we found that some moves were more characteristic of one phase than the other. Specifically:

1. A higher percentage of experts' dialogue turns during PPR than during problem solving contained at least one positive feedback statement, or *Appraise* act ($\chi^2 = 18.00$; $p < .001$). The same distinction applied to explanations (*Explain* acts; $\chi^2 = 11.13$, $p < .001$).
2. Particular types of explanations were more prevalent in experts' PPR contributions than in their problem-solving contributions—namely, explanations about how the system as a whole and its components work (*Knowledge Explanations;* $\chi^2 = 13.80$; $p < .001$); about the causal processes involved in transmitting a correct or faulty signal through a circuit (*Causal Process Descriptions*; $\chi^2 = 19.98$; $p < .001$), and about why to take an action in light of the results of prior actions (*Step Explanations*; $\chi^2 = 19.98$; $p < .001$).
3. A higher percentage of students' turns during PPR contained requests for explanations about system function and operation (that is, *Knowledge Explanations*) than during problem solving ($\chi^2 = 5.61$; $p < .001$).
4. Student self-explanations and self-appraisals were also much more prevalent during PPR than during problem solving ($\chi^2 = 27.05$; $p < .001$). Several studies (e.g., Chi, Bassok, Lewis, Reimann, and Glaser, 1989) have shown that self-explanations correlate strongly with learning.

Not surprisingly, some types of speech acts occurred more frequently during problem solving than during PPR, such as student requests for advice, feedback, and help with interpreting test results. Students seldom supported advice given to peers with explanations, either during problem solving or during PPR. When they did explain during problem solving, students tended to issue evidential explanations (*Step explanations*) that were unsupported by discussions of system function and operation ($\chi^2 = 11.10$; $p < .001$).

Taken together, the results reported above suggested that the principle role of PPR was to expose the rich domain knowledge that expert avionics technicians draw upon to guide their troubleshooting actions. During problem solving, students and their mentors may have been too focused on solving the task at hand to attend to the conceptual knowledge underlying strategic rules.

We extended this comparative analysis by investigating the rhetorical structure of experts' explanations during each instructional phase. This work was supported by machine learning techniques, as described in Katz, Aronis, and Creitz (1999). The machine learning software generated descriptions of the speech act structure and composition of *Knowledge Explanations* and illuminated instructional roles of speech acts that we could not have predicted. For example, it told us that *Identify* acts tend to occur near the beginning of a *Knowledge Explanation*. This cued us to look at the data that matched this pattern in order to try to explain it. We discovered that *Identify* acts do more than establish shared reference. Calling attention to a domain object's function-suggesting label (e.g., *Reset pin*), through an *Identify* act, was one of the mechanisms by which experts constructed explanations about how the electronic system investigated in **Sherlock 2** works (e.g., "*Pin 48 is a Reset pin. If it goes to 5 volts, it resets the test station.*"). Although the machine learning techniques that we developed were limited with respect to the scope of discourse structure that they could model, they demonstrated the promise of this approach for understanding how the structure of tutorial explanations supports specific instructional goals.

What role, if any, does reflective dialogue play in eliminating misconceptions?

Motivation

The differences that we observed between dialogue moves during and after practice exercises suggested that PPR provided a forum for discussions about conceptual issues that participants glossed over during problem solving. We therefore speculated that the reflective dialogues might have played a role in resolving some of the misconceptions that students revealed about how the system they were investigating works. Unfortunately, our comparative analysis of dialogue moves could help little with testing this hypothesis. It painted an impressionistic image of participants' rhetorical goals during each phase of instruction, but was devoid of detail about the semantic nature of these goals. Hence it raised the question, *were the observed differences significant for instruction?* For example, did the higher proportion of tutor-generated system knowledge explanations and student self-explanations during PPR predict whether misconceptions would be eliminated?

To analyze what role, if any, PPR plays in eliminating misconceptions, we shifted our focus to the instructional dialogue (ID) level of our representational scheme. Recall that at the ID level we encode semantic information such as the central strategic rule addressed by the exchanges forming the dialogue. At the contribution segment level, we encode other semantic information—e.g., the specific subrule or concept addressed in the segment, and whether the speaker has presented the rule or concept correctly (is the proposition true?). This information allowed us to track discussions about a misapplied rule across sessions. We could then determine if and when the mentor accurately addressed a misconception and whether the misconception was eliminated.

Methods

Hypotheses. We hypothesized that a misconception associated with a strategic rule was more likely to be eliminated by the next problem-solving task if discussion about the misconception was distributed between problem solving and PPR than if it took place only during problem solving. Our prediction was based on research showing that highly elaborated explanations during problem solving may be distracting or cause cognitive overload (Sweller, 1988; Sweller, van Merriënboer, & Paas, 1998). Hence, the dependent variable was repeated evidence of a misconception and the independent variable was staging of discussion in the previous exercise that required correct application of the rule.

We made no prediction as to whether distributed discussions about misconceptions would be more effective than discussions that took place only during PPR. Also, we did not expect that addressing a misconception would correlate with learning the strategic rule itself. A student can have multiple misconceptions about a rule which prevent him from applying it appropriately. He may also understand the conditions for rule "firing," but be unable to execute the actions correctly. Hence, our focus was on particular misconceptions and whether the staging of discussions about these misconceptions correlated with their recurrence.

Task Description. For each student, we identified a set of strategic rules that the student misapplied because of a misconception apparent in his actions or dialogue. We focused on the rules shown in Appendix 2, which also presents a sample of misconceptions that students have about these strategic rules. Knowledge of these rules is required in order to troubleshoot a particular type of circuit board called a "Test Point Select Relay Card." A model of this board and its surrounding circuitry is shown in Figure 2. Students typically understand that they need to check the outputs of a relay card like the A1A3A15 (pins 11 and 12), and—if the output reading is incorrect—that they should check all of the component's inputs. The card has two main inputs: (1) the inputs on the path between the aircraft unit under test (UUT) and the measurement device (pins 57 and 58), and (2) the inputs from the A2A3A10 logic card, which are responsible for activating *only one* relay in the series represented as B20-B29. However, sometimes something goes wrong and more than one relay is "told" to activate. When this happens, the signal can not pass through the relay card. For example, if B24 is the correctly activated relay in Figure 2, the

signal will be blocked if the logic signals coming from the A2A3A10 board incorrectly also set relay B23. Nothing will happen, though, if relay B25, for example, is activated, since this relay is not in the path between B24 and the relay board's outputs (pins 11 and 12).

The avionics students who participated in the study invariably reached an impasse when they found that they had incorrect outputs, even though "all" of the card's inputs were good—that is, the signal path inputs (at pins 57 and 58) and the inputs to the relay that should be activated (B24, in our example). In their faulty mental model of this circuitry, they did not realize that there are other control inputs to the relay board that can thwart its ability to route a signal—namely, the control signals to relays B20-B23. This is represented as misconception 4 in Appendix 2. Other misconceptions prevent students from realizing that the signal to the selected relay is good (misconceptions 1 and 2); from identifying the activated relay (misconception 3); from testing the unselected relays (misconceptions 5 and 6); and from testing the inputs to the logic card (misconceptions 7 and 8).

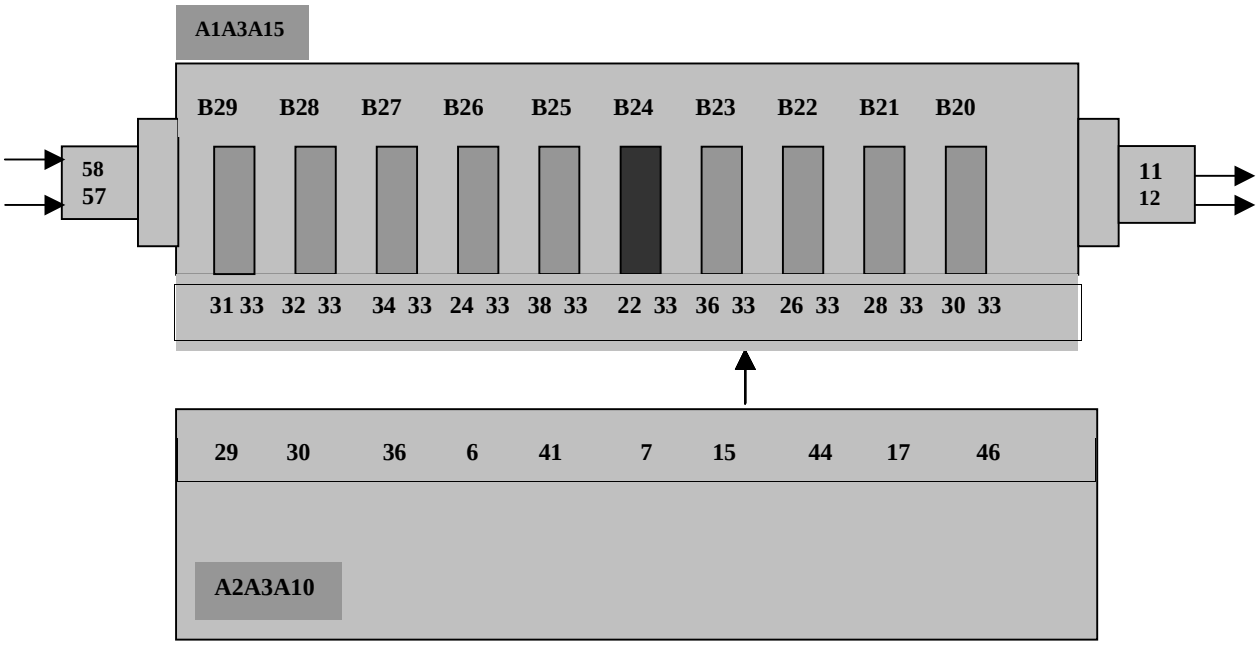

Figure 2. Abstract diagram of a circuit path

Analysis. Our unit of analysis was a student-rule pair. The corpus consisted of thirty student-rule pairs, involving the rules stated in Appendix 2. Twelve students who participated in the study are represented in this sample.

For each student-rule pair, we identified each instructional dialogue wherein the rule was applied or addressed. By "applied," we mean that the student spoke of taking the action specified in the rule or the mentor simply advised the student to carry out the action, without explaining why. By "addressed," we mean that the rule was explained. To restate our hypothesis in terms of this distinction: We predicted that ID's in which a strategic rule was applied would not correlate with elimination of a misconception about the rule. However, ID's in which a rule was addressed would correlate with elimination of a misconception, but only if the explanation contained statements that directly contradicted or corrected the misconception.

It was common for several ID's about a rule to occur during a problem-solving exercise. There could be one problem-solving ID about the rule and one PPR ID, several problem-solving ID's and no PPR ID's, etc. Since we were interested in all of the ID's that a student engaged in between one problem-solving exercise in which competency in a strategic rule was required and the next, our main linguistic unit of analysis was an *ID Series*, rather than a single ID. An ID Series can consist of a single ID or multiple ID's. Approximately 37% of ID Series contained one ID, 53% contained two ID's, and 10% contained three ID's. Table 3 illustrates an ID Series about Rule 51 (Appendix 2). One ID took place during problem solving, the other during PPR.

The solver has the misconception that the only relevant relay in the path is the active relay (misconception 4). The mentor briefly addresses this misconception in the problem-solving ID (shown in italics), and elaborates in the post-practice ID.

In approximately 62% of the thirty student-rule cases, the student misapplied the rule repeatedly, across two or more exercises. Because rules were typically addressed during each exercise in which they were misapplied, the data corpus consisted of forty-nine ID Series. This was the size of the corpus after we excluded ID Series containing incorrect explanations, and those involving collaborating students who shared a misconception, if the ID Series was successful for one student but unsuccessful for the other.

For each ID Series, we coded the following features about the associated misconception:

- *Addressed:* Was the student's misconception about the strategic rule addressed during the ID Series?
- *Staging:* When was the misconception addressed—during problem solving only, during PPR only, or distributed across the two phases of instruction?
- *Effectiveness:* Was the misconception repeated in the next problem-solving session?

We coded whether the misconception demonstrated by the student had been addressed because, as stated previously, a rule can be applied or discussed without being addressed (explained), and a rule can be addressed without targeting the student's misconception. The latter case is common when the mentor provides a script-based explanation about a rule (Sleeman et al., 1989). A script-based explanation covers the main facts and concepts that the tutor thinks the student should know, based on the tutor's "curriculum script" (Putnam, 1987), without tailoring the explanation to particular misconceptions that a student may have. Two raters coded each ID Series for the features described above. Agreement on staging of instruction was 88% (kappa = .85); agreement on effectiveness was 73% (kappa = .62). Unreliable codes were discussed and re-coded.

Results

Students' misconceptions about the rules stated in Appendix 2 were addressed in 78% (38) of the forty-nine ID series. Staging of discussion about misconceptions occurred as follows: 27% took place during problem solving only, 29% during PPR only, and 44% were distributed between problem solving and PPR. In approximately 2/3 of the distributed cases, the PPR discussion elaborated upon the problem-solving discussion; in 1/3 of distributed discussions, an explanation was simply restated during PPR.

As shown in Table 4, addressed misconceptions were less likely to recur in the next problem that required the associated strategic rule than were unaddressed misconceptions (p(Fisher) = 0). Six of the forty-nine ID Series were eliminated from this analysis, because the status of the misconception could not be determined; hence N = 43 in Table 4. (Figures in parentheses represent percent of total.) At least in this task domain, the data suggest that not addressing a misconception almost guaranteed that it would resurface. The misconception had to be directly contradicted or corrected in order for it to be resolved. Approximately 76% (26/34) of the misconceptions that were addressed were not repeated. However, addressing a misconception did not guarantee that it would be resolved. Approximately 24% (8/34) of addressed misconceptions recurred.

Table 4. Relationship between addressing misconceptions and recurrence

	Repeated	**Not Repeated**
Not Addressed	9 (21%)	0 (0%)
Addressed	8 (19%)	26 (60%)
N = 43	p(Fisher) = 0	

Analysis of the relationship between staging of discussions about misconceptions and repetition of misconceptions is shown in Table 5. Among the thirty-four addressed misconceptions represented in Table 4, the staging of the discussion (problem solving only, PPR only, etc.) for two ID's could not be determined; hence N = 32 in Table 5. The results marginally support our hypothesis: *distributed discussion of misconceptions was more effective than instruction that took place only during problem solving* ($\chi^2 = 5.73$; $p = .057$). Although the data in Table 5 suggest that misconceptions addressed only during PPR were less likely to be repeated than those addressed only during problem solving, this relationship did not reach significance.

Table 5. Relationship between when misconceptions were addressed and recurrence

When Addressed	Repeated	Not Repeated
PPR only	2 (6%)	7 (22%)
Distributed	1 (3%)	14 (44%)
Problem solving only	4 (13%)	4 (13%)
N = 32	*p* = .057	

As mentioned previously, we did not expect that addressing a particular misconception about a strategic rule would correlate with correct application of the rule, because the student might have different types of misunderstanding about the rule: strategic, procedural, and/or conceptual. The analysis shown in Table 6 did not support this prediction. Among the forty-nine ID's, the status of the student's ability to apply the strategic rule could not be determined in one case; hence, N = 48 in Table 6. Addressing a misconception was correlated with correct rule application (*p*(Fisher) = .01). This suggests that, in the majority of cases, each misconception represented in Table 6 was the only error blocking correct application of the associated strategic rule.

Table 6. Relationship between addressing misconceptions and rule application

	Misapplied	Correctly Applied
Not Addressed	9 (19%)	2 (4%)
Addressed	12 (25%)	25 (52%)
N = 48	*p*(Fisher) = .01	

Recall that our comparative analysis of speech act frequencies during problem-solving and reflective dialogue revealed that some speech acts were more prevalent during PPR—e.g., tutor-generated system knowledge explanations, and student-generated self-explanations. Did the occurrence of these speech acts in ID's that addressed a misconception correlate with elimination of the misconception? Did the PPR discussions contribute to the elimination of misconceptions? The analysis shown in Table 7 suggests that self-explanation correlated with non-repetition of misconceptions (*p*(Fisher) = .05). In 93% (14/15) of ID's in which the student restated or elaborated upon the principle that the mentor explained, the misconception did not resurface. This analysis was conducted on the thirty-four ID's in which misconceptions were addressed and the status of the misconception could be determined.

Table 7. Relationship between self-explanations and recurrence of misconceptions

	Repeated	Not repeated
No self-explanation	7 (21%)	12 (35%)
Student self-explained	1 (3%)	14 (41%)
N = 34	*p*(Fisher) = .05	

Table 8 shows that most self-explanations (80%) took place during PPR or were distributed between PPR and problem solving. In 92% (11/12) of cases in which a self-explanation occurred solely during PPR or was distributed, the misconception was not repeated. However, the relationship between self-explanation during PPR—alone or distributed between PPR and problem solving—and recurrence of misconceptions was not significant (*p*(Fisher) = 1.0).

Table 8. Relationship between when self-explanation occurred and recurrence of misconceptions

	Repeated	**Not repeated**
Problem-solving and PPR	0	2 (13%)
PPR only	1 (7%)	9 (60%)
Problem solving only	0	3 (20%)
N = 15	*p* = .77	

Although Tables 7 and 8 suggest that self-explanation correlated with the elimination of misconceptions irrespective of staging, the analysis shown in Table 9 indicates that receiving a system knowledge explanation did not in itself predict that a misconception would be eliminated (*p*(Fisher) = .36). However, knowledge explanations that occurred during PPR only, or were distributed between PPR and problem solving, significantly predicted the non-recurrence of misconceptions (*p*(Fisher) = .005). This relation is shown in Table 10, with respect to the twenty-six ID's that contained knowledge explanations. Perhaps students were more likely to attend to explanations presented solely during PPR, or elaborated upon during PPR, than to those presented during problem solving.

Table 9. Relationship between knowledge explanations and recurrence of misconceptions

	Repeated	**Not repeated**
Student did not receive a knowledge explanation	3 (9%)	5 (15%)
Student received knowledge explanation	5 (15%)	21 (62%)
N = 34	*p*(Fisher) = .36	

Table 10. Relationship between when knowledge explanations occurred and recurrence of misconceptions

	Repeated	**Not repeated**
Knowledge explanation during problem solving only	4 (15%)	2 (8%)
Knowledge explanation during PPR only, or distributed	1 (4%)	19 (73%)
N = 26	*p*(Fisher) = .005	

In addition to yielding the above findings on the effects of presenting explanations at different times, the data corpus suggested a way of describing the macro-level structure of reflective dialogues. Typically, the mentor initiated these discussions. Students joined in to respond to their mentors' comments and questions and to bring new issues to the table. Sometimes mentors then initiated new discussions and the cycle continued. Two main factors distinguish the manner in which mentors presented their initiating turns: (1) whether it was *narrative*, providing students with a stepwise recap of their solution, with feedback, and (2) whether feedback was *error-centered* or gave balanced attention to the positive and negative

aspects of students' performance. Taken together, these features define four ways in which mentors conducted discussions about strategic rules and "mal-rules":

1. Narrative with mixed feedback
2. Narrative and error-centered
3. Non-narrative with mixed feedback
4. Non-narrative and error centered

The data is shown in Table 11. Overall, no strong preference is visible for either feature. 54% of mentors' initiations were non-narrative; 45% were narrative. 48% contained mixed feedback (positive and negative), while 51% were error-centered. However, the data in Table 11 suggest that four of the mentors conducted reflective discussions in a preferred manner. Mentors AK and BK always focused on errors, in non-narrative fashion; DB always walked students through their solution and pointed out its strengths and weaknesses, while DH focused on errors within a narrative structure. Because the number of PPR sessions that each mentor conducted was small (3-6), this analysis raises the need for further research on the macro structure of reflective dialogue, as well as evaluation of the four approaches listed above. The results of this research would inform the design of automated "mentors" to conduct PPR in intelligent tutoring systems.

Table 11. Mentors' approach to post-practice reflection

Approach	**Mentor**								
	AK	**BG**	**BK**	**DB**	**DC**	**DH**	**MD**	**PG**	**%Total**
Narrative, mixed feedback	0	0	0	100%	20%	20%	50%	33%	33%
Narrative, error centered	0	0	0	0	20%	60%	0	0	12%
Non-narrative, mixed feedback	0	33%	0	0	20%	20%	0	67%	15%
Non-narrative, error centered	100 %	67%	100 %	0	40%	0%	50%	0	39%
Sessions	4	3	3	6	5	5	4	3	33

DISCUSSION AND CONCLUSIONS

Although it makes intuitive sense that misconceptions need to be addressed in order to be resolved, previous research on tutoring does not support this notion. For example, Sleeman, Kelly, Martinak, Ward and Moore (1989) found that tutors who explained why algebra "mal-rules" were wrong were not more successful than tutors who merely pointed out the error and demonstrated the correct procedure. In their discussion of this finding, the authors emphasized the need to replicate their research in different task domains, with different students, etc. Research on tutoring in physics by VanLehn, Siler, Murray, and Baggett (1998) suggests that misconceptions about some principles must be "untaught;" they can not simply be overridden. For example, students who applied a particular kinematics equation incorrectly and received an explanation about their error did better on similar post-test problems than students who were simply told that their equation was wrong. Nonetheless, for the majority of rules that VanLehn et al. (1998) analyzed, there was no significant correlation between addressing misconceptions and learning.

Hence, the current study adds to the confusion about whether or not addressing misconceptions helps to resolve them. Perhaps the positive correlation we found can be attributed to the difference in task domain: maybe misconceptions need to be addressed in diagnostic tasks like electronic troubleshooting, where evidential reasoning is grounded in system understanding, but not in the procedural aspects of algebra or quantitative physics. This in turn suggests that deep student modeling would be more beneficial for teaching conceptual knowledge than for teaching procedural knowledge. Clearly, Sleeman et al.'s (1989) call for further research on this issue remains relevant.

There were three occasions in which mentors addressed misconceptions. First, and most commonly (80% of ID's), mentors responded to conceptual errors that were revealed through students' dialogue and/or actions. In nearly all of these cases, some verbal evidence of the misconception accompanied students' faulty actions. Students expressed faulty beliefs to their mentor, while seeking advice about how to coach their peer. They also revealed misconceptions while negotiating the next action with their peer (Katz & O'Donnell, 1999). This shows that one advantage of studying educational dialogue in collaborative learning situations, as opposed to one-on-one tutoring, is that the former provides higher resolution on students' beliefs. It is debatable, though, whether this advantage offsets the disadvantage of increased inconsistency in the data—e.g., an explanation resolves a misconception for one student, but not the other.

The second way that students' misconceptions became addressed was through script-based explanations about the strategic rule. In these cases (14% of ID's), the mentor gave a general explanation about the strategic rule which was part of his agenda, or "curriculum script" (Putnam, 1987). Script-based explanations were typically provided in response to an error that did not reveal a particular misconception; the student misapplied or showed confusion about the rule, but it was unclear why. The mentor responded, it seemed, by "covering all bases." Perhaps serendipitously, the script-based explanation included information that addressed a misconception that the student had.

The third and least common occasion for addressing misconceptions was through cognitive diagnosis. In only 6% of cases, tutors explicitly probed students to explain their error. This observation is consistent with previous research, which shows that tutors rarely do detailed diagnosis before remediation (McArthur, Stasz, & Zmuidzinas, 1989; Putnam, 1987). In the present study, there may have been even less need for tutors to diagnose errors than during one-on-one interactions, because of the larger window into students' misconceptions that collaborative dialogue provides. When it did happen, mentors initiated cognitive diagnosis through a challenge move such as, "Why did you test pin 57?," or a statement of puzzlement.

As stated previously, in approximately 24% of cases, an addressed misconception was repeated. The number of cases (8) was too small to reveal a pattern, but a few cases suggested possible explanations. One obvious explanation is that a misconception can be so strongly ingrained that one episode of instruction is insufficient to eliminate it. This seemed to be the case for one student, who received nearly the same explanation about misconception 6 in Appendix 2 (*Pin numbers on the logic card correspond to relay numbers.*) several times before it was resolved. A second possibility is that students need to experience an impasse stemming from a misconception in order for an explanation to be meaningful and effective (VanLehn, Jones, and Chi, 1992). As discussed above, in script-based explanations about a strategic rule, the tutor sometimes *anticipates* misconceptions that could prevent correct application of a rule. However, this usurps the student's opportunity to reach an impasse.

This study provides empirical support for including post-practice reflection in ITS's and instruction in general. It reinforces Smith-Jentsch et al.'s (1996) finding that effective debrief correlates with effective performance. Our analyses suggest that it is better to address misconceptions during and after practice exercises than only during problem solving. It is important to note, however, that this is a correlational finding. Further research is needed to understand the causal mechanisms at play. Perhaps the effect can be explained in terms of cognitive load theory, as we originally proposed. As research by Sweller and others suggests (e.g., Sweller, van Merriënboer, & Paas, 1998) "less is best" in learning situations. It may be that students are too busy during problem solving to process hints and explanations that address misconceptions. During PPR, on the other hand, the student's mind is free to attend to these

explanations. This interpretation is supported by our finding that certain speech acts that were more prevalent during PPR than during problem solving, such as self-explanation by students and tutors' system knowledge explanations, predicted the elimination of misconceptions.

An alternative explanation for the relation between PPR and learning that we observed is that some of the hints and explanations that the mentors in this study issued during problem solving were too vague and non-directive to be effective. For example, in Table 3, the mentor hinted that there are other relays between the selected relay and the card's outputs. Perhaps this hint was cryptic without the elaboration provided by the PPR ID. It is also possible that mentors would have provided more elaboration during problem solving if they had been speaking rather than typing, and that the effect we observed was partly an artifact of the experimental setup.[ii]

Prior research on tutoring suggests that highly elaborated, didactic explanations do not correlate with learning (Chi, 1996). For example, VanLehn et al. (1998) found that shorter explanations worked best, with respect to learning a set of physics principles. In view of this research, it is surprising that, in the current study, distributed explanations were more effective than explanations that took place during one phase (problem solving or PPR only). PPR explanations such as the one shown in Table 3 tended to be lengthy. However, in prior research investigating length as a variable, explanations took place entirely during problem solving. Our analysis suggests that elaborated explanations may support learning (or at least not inhibit it), if they are parceled between problem solving and PPR. Further research is needed to test this hypothesis, in various domains.

We close with some remarks on the implications of this study for educational dialogue analysis. First, this study underscores the importance of attending to multiple aspects of discourse: syntax, semantics, and pragmatics. Our initial comparisons of rhetorical moves (speech acts) in problem solving and PPR had limited value. This analysis suggested that participants' intentions—as reflected in their speech acts—were different during the two phases, but did not disclose the instructional impact of these differences. It was not until we supplemented this pragmatic analysis with an examination of the semantic content of rhetorical moves that we could investigate the instructional effect of staging instructional discussions in different ways. Secondly, this study suggests that analyses of the relationship between instructional dialogue and learning require student-tutor interactions that span several sessions. [Sleeman et al. (1989) make a similar point.] Short sessions—eg., twenty minutes in Putnam's study (1987); fifty minutes in Sleeman et al.'s (1989) study—are common in tutoring research. As we have noted, we could only explain why some performance errors were committed when the underlying misconception re-surfaced in a later session. Also, in some cases it was not clear that a misconception had been eliminated until several sessions *after* an explanation had been provided.

Stated generally, the goals of educational dialogue analysis are to identify the features that distinguish instructional discourse from other types of discourse, and to determine what makes it effective. We view the research discussed in this paper as a first step towards understanding how reflective dialogue differs from dialogue during problem solving, and how the two phases of instruction work together to produce learning.

ACKNOWLEDGMENTS

This research was supported by grants from the Spencer Foundation and the Office of Naval Research, Cognitive Science Division (grant number N00014-97-1-0848). The data presented, the statements made, and the views expressed are not necessarily endorsed by the funding agencies. The authors thank David Allbritton, John Aronis, and Colin Creitz for assistance with data analysis, and Amy Soller and the reviewers for helpful comments on a previous version of the paper.

REFERENCES

Bielaczyc, K., Pirolli, P., & Brown, A. L. (1993). Strategy training in self-explanation and self-regulation strategies for learning computer programming. Report No. CSM-5. Berkeley, CA: University of California, School of Education.

Bull, S., & Broady, E. (1997, August). Spontaneous peer tutoring from sharing student models. In B. du Boulay and R. Mizoguchi (Eds.), *Proceedings of AI-ED 97: World Conference on Artificial Intelligence in Education*, Kobe, Japan. Ohmsha: IOS Press.

Chandrasekaran, B., Tanner, M. C., & Josephson, J. R. (1989). Explaining control strategies in problem solving. *IEEE Expert, 4(1),* 9-24.

Chi, M. T. H. (1996). Constructing self-explanations and scaffolded explanations in tutoring. *Applied Cognitive Psychology, 10,* 1-17.

Chi, M. T. H., Bassok, M., Lewis, M., Reimann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science, 13,* 145-182.

Chi, M. T. H., & VanLehn, K. (1991). The content of physics self-explanations. *Journal of the Learning Sciences, 1,* 69-105.

Collins, A., & Brown, J. S. (1988). The computer as a tool for learning through reflection. In Mandl, H. and Lesgold, A. (eds.), *Learning Issues for Intelligent Tutoring Systems* (pp. 80-113). New York: Springer-Verlag.

Collins, A., Brown, J. S., & Newman, S. E. (1989). Cognitive apprenticeship: Teaching the crafts of reading, writing, and mathematics. In L.B. Resnick (Ed.), *Knowing, learning, and instruction: Essays in honor of Robert Glaser*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Di Eugenio, B., Moore, J. D., & Paolucci, M. (1997, July). Learning features that predict cue usage. *Proceedings of the 35th Conference of the Association for Computational Linguistics (ACL97),* Madrid, Spain.

Gott, S. P. (1990). A naturalistic study of transfer: Adaptive expertise in technical domains. Paper presented at the Annual Meeting of the American Educational Research Association, Boston, MA.

Katz, S. (1995). Identifying the support needed in computer-*supported* collaborative learning systems. In J. L. Schnase, and E. L. Cunnius (Eds.), *Proceedings of CSCL '95: The First International Conference on Computer Support for Collaborative Learning* (pp. 200-203), Bloomington, Indiana. New Jersey: Lawrence Erlbaum Associates.

Katz, S., Aronis, J., & Creitz, C. (1999). Modeling pedagogical interactions with machine learning. In S. P. Lajoie and M. Vivet (Eds.). *Proceedings of the Ninth International Conference on Artificial Intelligence in Education*, Le Mans, France (pp. 543-50). Amsterdam: IOS Press.

Katz, S., Lesgold, A., Eggan, G., & Gordin, M. (1993). Modeling the student in Sherlock II. *Journal of Artificial Intelligence in Education, 3,* 495-518.

Katz, S., Lesgold, A., Hughes, E., Peters, D., Eggan, G., Gordin, M., Greenberg., L. (1998). Sherlock 2: An intelligent tutoring system built upon the *LRDC Tutor Framework*. In C. P. Bloom and R. B. Loftin (Eds.), *Facilitating the Development and Use of Interactive Learning Environments* (pp. 227-258). New Jersey: Lawrence Erlbaum Associates.

Katz, S., & O=Donnell. (in press). The cognitive skill of coaching collaboration. In C. Hoadley & J. Roschelle (Eds.), *Proceedings of Computer Support for Collaborative Learning (CSCL) 1999* (pp. 291-299), Stanford, CA.

Lederman, L. C. (1992). Debriefing: Toward a systematic assessment of theory and practice. *Simulation & Gaming, 23(2),* 145-160.

Lesgold, A., Eggan, G., Katz, S., & Rao, G. (1992). Possibilities for assessment using computer-based apprenticeship environments. W. Regian & V. Shute (Eds.), *Cognitive approaches to automated instruction* (pp. 49-80). Hillsdale, NJ: Lawrence Erlbaum Associates.

Mann, W., & Thompson, S. (1988). Rhetorical structure theory: Towards a functional theory of text organization. *TEXT, 8(3),* 243-281.

McArthur, D., Stasz, C., & Zmuidzinas, M. (1990). Tutoring techniques in algebra. *Cognition and Instruction, 7,* 197-244.

Moore, J. D. (1996). Making computer tutors more like humans. *Journal of Artificial Intelligence in Education, 7(2),* 181-214.

Moore, J. D., Lemaire, B., & Rosenblum, J. A. (1996). Discourse generation for instructional applications: Identifying and exploiting relevant prior explanations. *Journal of the Learning Sciences, 5(1),* 49-94.

Pilkington, R. M. (1997). Analyzing educational discourse: The DISCOUNT scheme. Computer Based Learning Unit, University of Leeds, *Technical Report No. 019703*.

Pioch, N. J., Roberts, B., & Zeltzer, D. (1997). A virtual environment for learning to pilot remotely operated vehicles. *Proceedings of Virtual Systems and Multimedia 1997,* Geneva, Switzerland.

Putnam, R. T. (1987). Structuring and adjusting content for students: A study of live and simulated tutoring of addition. *American Educational Research Journal, 24,* 13-48.

Roberts, B. (1993). Constrained learning environments for intelligent tutoring. In P. Brna, S. Ohlsson, & H. Pain (Eds.), *Proceedings of the World Conference on Artificial Intelligence in Education* (pp. 521-528), Edinburgh, Scotland.

Rosé, C. P. (1997). The role of natural language interaction in electronics troubleshooting. *Proceedings of the Eighth Annual International Energy Week Conference and Exhibition*.

Searle, J. (1969). *Speech Acts.* Cambridge, England: Cambridge University Press.

Sinclair, J. & Coulthard, M. (1992). Towards an analysis of discourse. In M. Coulthard (Ed.), *Advances in spoken discourse analysis* (pp. 1-34). London: Routledge.

Sleeman, D., Kelly, A. E., Marinak, R., Ward, R. D., & Moore, J. L. (1989). Studies of diagnosis and remediation with high school algebra students. *Cognitive Science (13),* 551-568.

Smith-Jentsch, K. A., Zeisig, R. L., Acton, B., & McPherson, J. A. (1998). Team dimensional training: A strategy for guided team self-correction. In J. A. Cannon-Bowers & E. Salas (Eds.), *Making decisions under stress: Implications for individual and team training* (pp. 271-297). Washington, DC: APA.

Stiles, W. B., & Putnam, S. M. (1992). Verbal exchanges in medical interviews: concepts and measurement. *Social Science and Medicine, 35(3),* 347-355.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning: *Cognitive Science, 12,* 257-285.

Sweller, J., van Merriënboer, J., & Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10(3),* 251-296.

VanLehn, K., Jones, R. M., & Chi, M. T. H. (1992). A model of the self-explanation effect. *Journal of the Learning Sciences, 2,* 1-59.

VanLehn, K., Siler, S., Murray, C., & Baggett, W. B. (1998). In M. A. Gernsbacher & S. Derry (Eds.), *Proceedings of the Twentieth Annual Conference of the Cognitive Science Society* (pp. 1084-9), Madison, Wisconsin. Hillsdale, NJ: Erlbaum.

APPENDIX 1: A SAMPLE CODED LEARNING CONVERSATION[1]

Student: Do I set my meter for Ohms or for volts? I am testing pins 14, 15 on the A2 Regulator.
Tutor: I suggest you backtrack and look at the relay card. You want to make sure the signal is getting to the relay card before going to the uut [sic].
Student: Ok, what is a "relay card?"
Tutor: a relay card is part of the measurement signal path. The schematics are located in the schematic book. Do you remember how to find the card you need?
Student: I look for the number, in this case 63, and look for the page in the schematics with B 63?
Tutor: close except you want A63
Student: Got it.

=== Instructional Dialogue 1 Starts ===
Focal Rule: when to test a TPS Relay Card
Approach: directive
Explicit: false
Staging: problem solving only
Rating: unsuccessful

<u>Contribution 1</u>
Speaker: student
Hearer: tutor
Text: *"Do I set my meter for Ohms or for volts? I am testing pins 14, 15 on the A2 Regulator."*

Contribution Segment 1.1: *Do I set my meter for Ohms or for volts?*
Speech Act: *Request* for advice-2
Predicate: Set-Test-Equipment--How (Meter)
Rating: neutral
Exchange#: 1
Role: initiation
Position: 1

Contribution Segment 1.2: *I am testing pins 14, 15 on the A2 Regulator.*
Speech Act: *Strategic justification* of request-1 for advice-2
Predicate: Verify-component (UUT card)
Rating: incorrect
Exchange#: 1
Role: initiation
Position: 2

***** Instructional Dialogue 2 Starts *****
Focal Rule: How to test a TPS Relay Card
Approach: guided
Explicit: true
Staging: problem solving only
Rating: successful

<u>Contribution 2</u>
Speaker: tutor
Hearer: student
Text: *"I suggest you backtrack and look at the relay card. You want to make sure the signal is getting to the relay card before going to the uut [sic]."*

Contribution Segment 2.1: *I suggest you backtrack and look at the relay card.*
Speech Act: Correct
Rule: When to test a TPS Relay Card
Subrule:
Conceptual Knowledge:
Predicate: Verify-component (TPS Relay Card)
Rating: partially correct
Exchange#: 1, 2
Role: response, initiation
Position: 1, 1

Contribution Segment 2.2: *You want to make sure the signal is getting to the relay card before going to the uut.*
Speech Act: *Strategic justification* of correction-1
Rule: When to test a TPS Relay Card

[1]Most abbreviations used for coding have been replaced by more readable codes. Subjects' references to each other by name have been deleted from the transcript.

Subrule:
Conceptual Knowledge:
Predicate: Verify-inputs (TPS Relay Card)
Rating: partially correct
Exchange#: 1, 2
Role: response, initiation
Position: 2, 2

Instructional Dialogue 2.1 Starts
Focal Rule: Concept of Arelay card@
Approach: directive
Explicit: true
Staging: problem solving only
Rating: indeterminate

<u>Contribution 3</u>
Speaker: student
Hearer: tutor
Text: *"Ok, what is a "relay card?"*

Contribution Segment 3.1: *Ok,.*
Speech Act: *Acknowledge* strategic-explanation-2 of correction-2
Predicate: Verify-component (TPS Relay Card)
Rating: neutral
Exchange#: 2
Role: response
Position: 1

=== **Instructional Dialogue 1 Ends** ===

Contribution Segment 3.2: *what is a "relay card"*
Speech Act: *Request* for definition-2
Predicate: Define-term (Relay Card)
Rating: neutral
Exchange#: 2, 3
Role: response, initiation
Position: 2, 1

<u>Contribution 4</u>
Speaker: tutor
Hearer: student
Text: *"a relay card is part of the measurement signal path. The schematics are located in the schematic book. Do you remember how to find the card you need?"*

Contribution Segment 4.1: *A relay card is part of the measurement signal path. The schematics are located in the schematic book.*
Speech Act: Hint
Rule: How to test a Test Point Select Relay Card
Subrule: How to find the location of the Test Point Select Relay Card in the schematics
Conceptual Knowledge: Concept of "relay card"
Rating: correct/incomplete
Exchange#: 3
Role: response
Position: 1

Contribution Subsegment 4.1.1: *A relay card is part of the measurement signal path.*
Speech Act: Inform
Predicate: Location (Relay Card)
Rating: correct

Contribution Subsegment 4.1.2: *The schematics are located in the schematic book.*
Speech Act: Inform
Predicate: Location (schematics)
Rating: correct

Instructional Dialogue 2.1 Ends

^^ **Instructional Dialogue 2.2 Starts** ^^
Focal Rule: How to find the location of the Test Point Select Relay Card in the schematics
Approach: guided
Explicit: true
Staging: problem solving only
Rating: successful

Contribution Segment 4.2: *Do you remember how to find the card you need?*
Speech Act: Diagnostic-query
Rule: How to test a TPS Relay Card
Subrule: How to find the location of the Test Point

Select Relay Card in the schematics
Conceptual Knowledge:
Predicate: Find-component-how (TPS Relay Card)
Rating: neutral
Exchange#: 3, 4
Role: response, initiation
Position: 2, 1

<u>Contribution 5</u>
Speaker: student
Hearer: tutor
Text: *"I look for the number, in this case 63, and look for the page in the schematics with B 63?"*

Contribution Segment 5.1: *I look for the number, in this case 63, and look for the page in the schematics with B 63?*
Speech Act: Explain
Predicate: Find-component-how (TPS Relay Card)
Rating: partially correct
Exchange#: 4
Role: response
Position: 1

Contribution Segment 5.2: *I look for the number, in this case 63, and look for the page in the schematics with B 63?*
Speech Act: *Request* for appraisal-2 of explanation-1
Predicate: Find-component-how (TPS Relay Card)
Rating: neutral
Exchange#: 4, 5
Role: response, initiation
Position: 1, 1

<u>Contribution 6</u>
Speaker: tutor
Hearer: student
Text: *"close except you want A63"*

Contribution Segment 6.1: *close except you want A63*
Speech Act: *Correct* explanation-2
Rule: How to test a TPS Relay Card
Subrule: How to find the location of the Test Point
Select Relay Card in the schematics
Conceptual Knowledge:
Predicate: Identify-relay (relay)
Rating: correct
Exchange#: 5, 6
Role: response, initiation
Position: 1, 1

<u>Contribution 7</u>
Speaker: student
Hearer: tutor
Text: *"Got it."*

Contribution Segment 7.1: *Got it*
Speech Act: *Acknowledge* correction-2 of explanation-1
Predicate: Find-component-how (TPS Relay Card)
Rating: neutral
Exchange#: 6
Role: response
Position: 1

^^ **Instructional Dialogue 2.2 Ends** ^^

*** **Instructional Dialogue 2 Ends** ***

APPENDIX 2: STRATEGIC RULES AND ASSOCIATED MISCONCEPTIONS

Strategic Rule	Misconception	Correct Principle
Rule 51: If the signal inputs to a test point select relay board are good, the inputs to the selected relay are good, but the board's outputs are bad, then verify the control signals to the relays between the selected relay and the board's outputs.	1. A bad output signals means that the relay is not being energized because it is malfunctioning.	A bad output signal means that the relay is malfunctioning, or the signal is being blocked by an incorrectly activated relay.
	2. A resistance reading on the activated relay (high to low side) should be low.	A resistance reading on the activated relay (high to low side) should be high, because the diode is not connected to anything.
	3. There is no connection between the diagnostic test procedures and verification of a test point select relay card.	The diagnostic test procedures indicate which relay should be activated.
	4. The only relevant relay on a test point select relay card is the selected relay.	All relays between the selected relay and the relay board's outputs are relevant.
Rule 58: To verify the control signals in the path between a selected relay and a relay card's outputs, do a voltage check high to low on all unselected relays in this path; they should read 0 VDC.	5. The reading on the unselected relays should be the same as on the selected relay.	The unselected relays should read 0 VDC; the selected relay should read 28 VDC, thus showing a voltage drop.
	6. Pin numbers on a logic card correspond to relay numbers.	The numbering of pins on relays is arbitrary.
Rule 60: If the outputs of a logic board that controls the selection of relays are good, then test the inputs to this logic board.	7. The input signals from a switch to the logic board are represented in binary.	The inputs from a switch to the logic board are represented as Binary Coded Decimal.
	8. If a logic card's outputs are bad, the logic card itself is bad.	If a logic card's outputs are bad, either the card is bad or its inputs are bad.

ENDNOTES

[i] All but two participants were male. Hence, we use the masculine pronoun throughout.

[ii] Recently, we conducted a follow-up study of one-on-one tutoring between avionics experts and students, also set in **Sherlock 2**. One of our goals was to determine if the results of the present study would hold given student-mentor interaction that is spoken during problem solving and PPR. The study involved four student-mentor pairs, rather than collaborating students and a mentor. (We were unable to recruit enough subjects to replicate the collaborative learning situation.) All four mentors participated in the initial study described in this paper. Participants spoke across a room divider. We avoided face-to-face interaction in order to alleviate the problem of ambiguous references that occur when participants point to schematics.

Cursory analysis of the data suggests that elaborated explanations were more common during problem solving in the all-spoken setting (recent study) than in the typed setting (initial study). Correspondingly, there was, overall, less discussion during PPR in the all-spoken setting. However, it is possible that these differences stemmed from the fact that students were working on their own rather than collaborating with a peer, and did not result from the difference in interaction medium during problem solving (speaking versus typing). For example, mentors may have needed to explain more during problem solving in the recent study because—unlike

the initial study—there was no peer present to provide advice and explanations. Hence, although the recent study supports the idea that typed interaction inhibited elaborated explanation during problem solving in the initial study, a direct comparison, involving collaborating students and more subjects, would be needed to conclude this.