The Cognitive Skill of Coaching Collaboration

Sandra Katz, Gabriel O'Donnell

Learning Research and Development Center, University of Pittsburgh

Abstract: We observed eight experienced avionics technicians as they coached collaborating dyads who worked on problems in **Sherlock 2**, an intelligent tutoring system for avionics. Data analysis focused on: (1) defining the cognitive skill of coaching collaboration, (2) identifying cues that coaches use to detect and diagnose peer interaction impasses, and (3) specifying how coaches remedy these impasses. Coaching collaboration is an extension of the task of mentoring in one-on-one interactions. It involves three levels of diagnosis: diagnosing problems in the task situation, in students' knowledge and problem-solving process, and in peer interactions. One-on-one instructional interactions–e.g., medical rounds consultations between an expert physician and a resident–involve only the first two levels. We describe cues that can signal to a human or automated coach that intervention is necessary and can indicate what the cause of a peer interaction impasse is. During problem solving, the avionics experts tended to focus on giving advice that would keep students on a productive solution path. They seldom addressed problems at the interaction level– e.g., by prompting a student who evidently knew what to do next to help his peer. However, experts and students used the post-practice reflective discussions to more fully address knowledge gaps, misconceptions, and interaction-level problems.

Keywords: discourse analysis of collaborative learning, qualitative analysis of collaborative learning

Introduction

In order to develop automated coaches that can support students during collaborative problem-solving exercises, we must first specify the types of situations in which collaborative learners need help, and how human coaches intervene during these situations. This knowledge can be gleaned by observing naturalistic *human*-supported collaborative learning interactions. This paper reports on the results of one such study.

We observed eight experienced avionics technicians as they coached collaborating pairs of students (dyads) while the latter worked on problems in **Sherlock 2**, an intelligent tutoring system for avionics (Katz, 1995; Katz et al., 1998). One student primarily worked on a problem while the other student assumed a helping role; then they reversed roles for the next problem-solving session. (We refer to these roles as "solver" and "student coach," respectively.) The avionics expert provided help when the student coach solicited advice, or when the expert deemed it necessary. (We refer to the expert as the "coach.") Interaction was typed using a chat facility. After the problem was solved, the

expert reviewed students' performance. This "debrief" was conducted orally. The corpus consists of approximately thirty-five coached collaborative sessions.

This paper reports on the results of an analysis of the cognitive skill of coaching collaboration, based on a sample of this corpus. Our findings have implications for designing automated coaches that can detect, diagnose, and remedy impasses during CSCL sessions.

The multiple roles of diagnosis in coaching collaboration

Stated generally, diagnosis involves formulating, testing, and revising hypotheses in order to explain an abnormal condition in a system under investigation–e.g., a patient undergoing a medical examination, a grounded aircraft undergoing troubleshooting. Gadd (1995) has proposed a *dual-diagnostic theory* of expert understanding in one-on-one consultation situations, such as rounds dialogues between an expert physician and a resident. According to this theory, the expert physician simultaneously performs two levels of diagnosis while the less experienced physician (the "presenter/investigator") describes a patient's condition and the results of diagnostic tests performed on the patient. At one level, the expert physician uses the information provided by the presenter/investigator to construct his or her own model of the patient's condition. Gadd (1995) refers to this as "patient-level diagnosis." At a second level, the expert evaluates the presenter/investigator's reasoning about the patient. Gadd (1995) refers to this diagnostic task as "presenter/investigator level diagnosis." While the *content* of the information that the resident presents is central to patient-level diagnosis, it is the presence or absence of that information, the order in which it was gathered, and how the resident interpreted it that is central to presenter/investigator level diagnosis.

During coached collaborative problem-solving situations, a human coach carries out the same diagnostic processes as in this one-on-one consultation situation. (For the sake of generality, we will refer to these processes as *task-level diagnosis* and *student-level diagnosis* instead of patient-level and presenter/investigator level diagnosis,respectively.) However, there are important differences between these two situations, namely: (1) there is more than one student in a collaborative problem-solving situation and (2) the coach can observe students' actions and dialogue and interact with them *while* they are solving the problem, not just afterwards. For task-level diagnosis, these differences in instructional format mean that the coach's model of the problem state can be constructed on line, as he or she observes students' actions and the results of those actions. For student-level diagnosis, these differences have positive and negative implications. On the up side, the coach can use students' dialogue as input to a dynamically constructed model of student ability, in conjunction with students' actions. On the down side, formulating a cognitive model of two or more students is far more complex than formulating a cognitive model of one student, challenging itself. Collaboration introduces novel ambiguities into student-level diagnosis. For example, what does it mean when the solver takes an inappropriate action and the student coach observes silently? Is the student coach tacitly accepting the solver's action, opposed but reluctant to say so, not paying attention, or confused?

But the main difference between one-on-one instructional interactions and coaching collaboration is that the latter requires the coach to perform a third level of diagnosis, which we call *interaction-level diagnosis*. We define this level by referring to two-student (dyadic) collaboration. When students reach an impasse, the coach is challenged to determine why. Do both students lack a critical piece of knowledge or share a misconception? Or is the knowledge deficit one-sided, but some other factor is preventing the "more knowledgeable peer" from resolving the problem–for example, students are referring to different domain objects that have the same or similar names; the student coach is not paying attention or is reluctant to criticize his peer? Interaction-level diagnosis depends on student-level diagnosis for credit/blame assignment–that is, to determine who has the cognitive deficit: one student, both, or neither, in the case of simple errors or "slips." If the cognitive deficit is shared, interaction-level diagnosis aligns with student-level diagnosis. The coach can conclude that the impasse is due to a shared cognitive deficit and intervene accordingly. However, if there is evidence that the deficit is one-sided, then the coach's task at the interaction level is to formulate a hypothesis about why the more knowledgeable student (at least in the current situation) is unable to resolve the impasse, test that hypothesis by examining cues in students' dialogue and actions, revise the hypothesis and posit new ones as necessary.

In sum, we propose a *triple-diagnostic theory* of expertise in coaching collaborative problem solving, particularly in domains that involve the formulation and testing of hypotheses to explain a situation. This theory extends Gadd's (1995) *dual-diagnostic theory* of one-on-one instructional consultation. In this paper, we focus on interaction-level diagnosis during dyadic collaboration. Like the first two levels of diagnosis, this layer is framed by two other instructional tasks: *detecting* that a problem exists (in this case, an interaction impasse) and *remedying* the problem. There are some general signs that students have reached an interaction impasse–for example, the student coach directly solicited advice from the expert/coach; students have discussed the problem state for awhile but have taken few or no productive actions. Detection stops and diagnosis begins when the coach attempts to explain these general signs of trouble: a shared knowledge deficit, miscommunication, lack of confidence on the part of the student coach, some combination of the above, etc. During remediation, the coach addresses the problem, possibly tailoring his intervention to the specific nature of the impasse. In the next section, we describe the study that led to the formulation of this theory and discuss our findings.

A study of how human coaches detect, diagnose, and remedy interaction impasses

Methodology

We analyzed a sample of the corpus of tutorial dialogues described in the Introduction. In particular, we selected the three problems that domain experts deemed the most difficult–problems that require students to troubleshoot digital logic circuitry–since we expected that these problems would cause students to reach impasses. We coded the transcripts of six coached dyads for these problems, yielding a total of eighteen coded transcripts. Our goal was to describe types of interaction impasses, linguistic and other behavioral cues

that allowed the coaches to detect and diagnose them, and the coaches' approach to remediating these impasses.

As discussed in Katz (1995), the data collection procedure was designed to highlight peer coaching impasses (PCI's). We directed students to try to help each other before soliciting advice from the human expert or automated coach. We also asked coaches not to intervene unless they felt that students were "going down the garden path." For the most part, participants followed these instructions.

Our unit of analysis was a section of a transcript that contains one peer coaching impasse and the domain expert's intervention. We define a *peer coaching impasse* as a situation in which a student solicits help from the coach in order to advise his peer or the coach intervenes voluntarily. We did not code as PCI's situations in which the coach intervened gratuitously–that is, where there was no evidence that students were having trouble resolving a problem. In order to capture the cues that might have led coaches to intervene voluntarily, we specified the range of the unit of analysis as starting from the last productive action taken and ending with the last statement of the coach's intervention. Among the eighteen transcripts, we identified and analyzed thirty-six interaction impasses.

For each PCI, we recorded the following information:

- *Detection cues.* A description of general features of the situation which might have signaled to the coach that students reached (or were reaching) an impasse. Our goal here was to identify cues that automated tutors could use, as well, to determine that an intervention is necessary.
- *Impasse diagnosis.* A description of the nature of the interaction impasse, in terms of credit/blame assignment–that is, who had the cognitive deficit preventing progress: the solver, student coach, both, or neither? If only one student held a cognitive deficit in a given situation, why was his peer ineffective in helping him? Our goal here was to categorize PCI's, and to determine the relative frequency of each type.
- *Diagnosis cues.* A description of features of the interaction that enabled us to diagnose the impasse. We assume that these cues also enabled coaches to diagnose impasses, if and when they attempted to do so, and could similarly serve an automated CSCL coach.

- *Approach to remediation.* We focused on three aspects of coaches' interventions:

- *Directness*: Did the coach issue direct, explicit advice or did he guide students in determining the next appropriate action on their own?
- *Staging of instruction*: When did coaches address student-level problems–that is, knowledge gaps and misconceptions? With what frequency were instructional explanations issued during problem solving, the post-practice review, or parceled across the two stages?

- *Addressing interaction level problems.* To what extent did coaches address problems in peer interaction? When did most interaction-level discussions occur–during problem solving or debrief?

Two raters coded these features of the thirty-six PCI transcript segments. Overall agreement was 88% (kappa = .75). Reliability for individual categories will be stated in the appropriate sections.

Results

Why do coaches intervene? Approximately 42% of interventions were solicited; 58% were voluntary. Rater agreement on type of intervention was 96% (kappa = .92).

We identified several features of PCI's that might have signaled to experts/coaches that there was a problem, and hence a need for unsolicited intervention. These features and their frequencies are shown in Table 1. The most common PCI detection cue was student dialogue with few (or no) productive actions. A significant time lapse with no dialogue, and the student coach's failure to respond to the solver's inappropriate actions, also coincided with coaches' interventions. These cues may co-occur within a single PCI.

What types of PCI's are there, and how are they detected? Although the cues in Table 1 can suggest that a PCI has occurred, they are of limited use in determining what the nature of the impasse is. In particular, they are too ambiguous for credit/blame assignment. For example, if students have fallen silent and no productive actions have been taken, we do not know if the student coach is as clueless as the solver, has a suggestion in mind but is letting his peer flounder for awhile, or had an attention lapse. Fortunately, each type of PCI has several "signature" features that serve to distinguish it from other types.

Table 2 summarizes the main types of PCI's, their relative frequency in the corpus, and the relationship between intervention type and PCI type. The numbers in parentheses (within the Frequency column) indicate the percent of dyads who experienced each type of impasse. Rater agreement on impasse diagnosis was 73% (kappa = .62). Table 2 shows that shared knowledge deficits were the most common type of PCI (39%), although one-sided deficits were almost as common (36%). Although it appears that solicited interventions were more common for shared knowledge deficits than for one-sided

deficits–and, conversely, that unsolicited interventions were more common for one-sided knowledge deficits–this relationship was not significant.
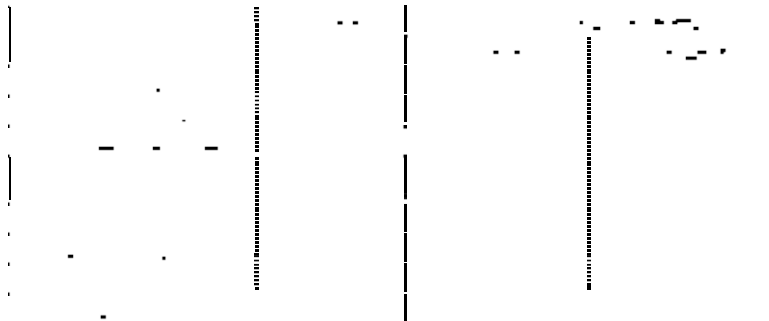
Taken together, shared and one-sided knowledge deficits represent 75% of the analyzed PCI's. We therefore focused on identifying "signature cues" for these types of impasses. Table 3 specifies these cues and their frequency. Note that diagnostic features can co-occur in a given PCI. Rater agreement on the presence of diagnostic features within PCI's, for all impasse types named in Table 2, was 86% (kappa = .67).

Shared knowledge deficits. There were several indicators that students had a knowledge deficit in common. As shown in Table 3, incorrect claims from both students about the current problem state–e.g., which strategic principle to apply, what a test result means– was the most common feature of this type of PCI. Other common indicators of shared knowledge deficits include: a student coach's overt or tacit acceptance of his peer's incorrect claims, faulty advice, and the need to solicit advice from the expert.

One-sided knowledge deficits. In a one-sided knowledge deficit, there is evidence that one student has the knowledge necessary to figure out what to do next, but the other student does not. The question then becomes, *why can't (doesn't) the more capable student help his peer?* We identified three main, potentially co-occuring–causes of peer coaching failures. As shown in Table 2, the most common cause is a breakdown of communication between students. Several PCI's were caused by unclear or ambiguous language. In the testing system simulated in **Sherlock 2**, many electronic components have the same pin numbers–e.g., there is a "pin 46" on two adjacent circuit boards. A telltale sign of referential ambiguity is a lot of talk intermixed with inappropriate actions and unsuccessful attempts by one student to clarify the correct referent; e.g., phrases such as "I'm talking about pin 46 on the A15 card."
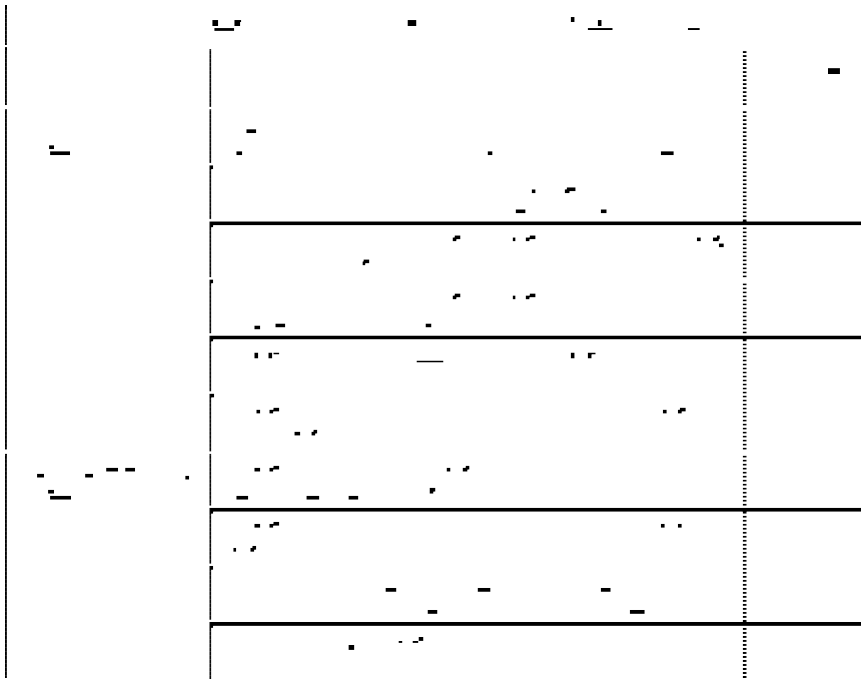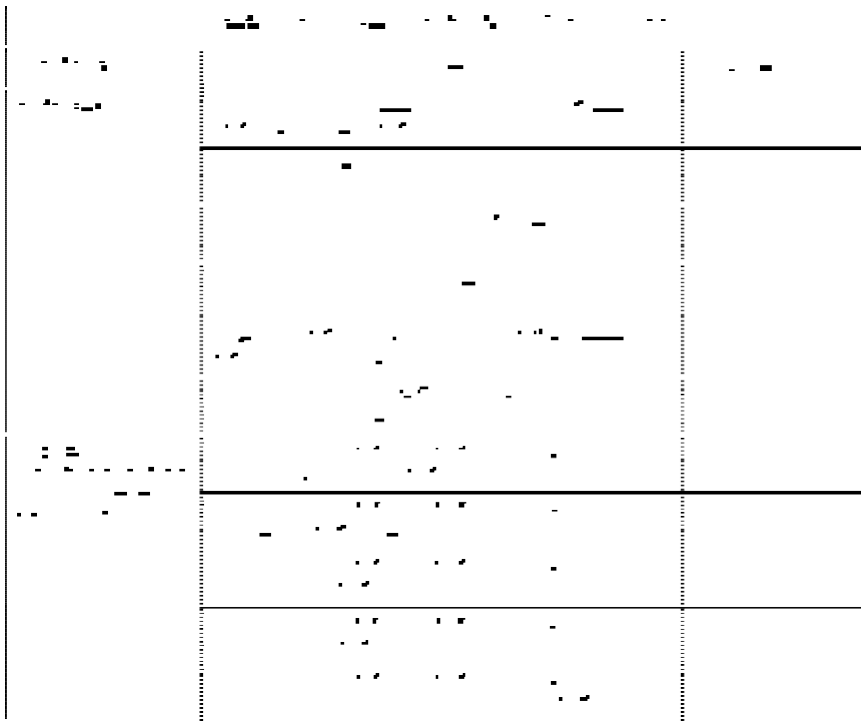
Another cause of peer coaching failures was poor pedagogy. Although unclear, vague, or ambiguous language on the part of a student coach can be viewed as a pedagogical problem, we treated communication breakdowns separately because they were the most common type of PCI involving one-sided cognitive deficits. Other pedagogical problems include:

- *Failure to intervene in a timely manner.* Some student coaches (some experts as well) didn't seem to know when it was time to step in. They let the solver flounder, even when it became obvious that he wasn't going to "get it" any time soon.

- *Misdiagnosis of peer errors, or unresponsiveness.* Students sometimes tried to diagnose their peer's misconception, but *mis*diagnosed it. At other times, the solver demonstrated a misconception but his peer ignored it.
- *Insufficiently directive coaching.* Some student coaches tried to imitate the expert/coach. They gave their peer hints in order to guide the latter in figuring out the next appropriate action. Unlike some experts who use hinting (Hume *et al.*, 1996), student coaches sometimes failed to make hints increasingly directive when the solver didn't "take the hint." They simply reworded the same hint, at the same information level, over and over.

Besides being rooted in communication and pedagogical problems, PCI's can stem from social factors. Correcting a peer was uncomfortable for some students. Table 4 illustrates a student coach's report during debrief (right side of table) about his uneasiness during problem solving telling the solver that he had overlooked a failed diagnostic test (left side of table). The fingerprint of uneasiness about critiquing is something we call *vicarious coaching*. During problem solving, the student coach addressed his comments to the expert, all the while aware that the solver would eventually see the student coach's interaction with the expert in the chat window. What distinguishes vicarious coaching from solicited help due to a shared knowledge gap is that the student coach's claims and ideas about how to advise his peer are correct in the former and incorrect in the latter. The prominent communicative act during vicarious coaching is the student coach's requests for the expert to provide feedback on his assessment of the solver's actions (e.g., Table 4 [1, 3, 5]), and on his plans for advising his peer (Table 4 [7])].

Parallel knowledge deficits. As shown in Table 2, 8% of the transcript segments contained what we call *parallel knowledge deficits*. In this case, students have separate, though possibly related knowledge gaps and misconceptions. They are signaled by students asking the coach different questions in a close timeframe, or by students making inappropriate claims about different domain concepts or strategic principles while interacting with each other. The coach has a full cognitive load when this happens. He is challenged to diagnose distinct problems at the student level, and to decide whether to address one or both problems.

Slips. In another 6% of PCI's, neither student had a cognitive deficit. The problem was simply due to a slip; e.g., performing a test on the A1A3A10 card, instead of the A2A3A10 card. Student coaches sometimes missed these seemingly minor errors, which had the potential to send students down the garden path (and often did).

As noted in Table 2, an additional 11% of the PCI's were coded as "Indeterminate," due to insufficient evidence for assigning credit and blame.

How did coaches remedy peer coaching impasses? Like classroom instruction, coaching dyadic or small group collaboration is a problem-solving task (Leinhardt & Greeno, 1986). We have discussed the diagnostic demands placed on the coach at the student and interaction levels. We turn now to several decisions coaches face about if, when, and how to address problems at both levels.

Interventions to address students' knowledge gaps and misconceptions. Table 5 summarizes our analysis of two aspects of coaches' interventions at the student level: directness of feedback and the staging of instructional explanations. In the majority of cases (67%), the coach issued direct advice or correction of a student's error during problem solving. 56% of interventions either relied solely on indirect approaches–e.g., hinting, prompting students to consider the next action or to reinterpret a result, referring students to Sherlock's advice–or combined indirect with direct feedback. Rater agreement on directness was 92% (kappa = .75). There was no correlation between directness of feedback and type of impasse (shared, one-sided, etc.).

Although coaches seemed to focus on keeping students on a productive path during problem solving, coaches and students alike turned their attention more towards *understanding* during the post-practice discussions. As shown in Table 5 (Staging of Instruction…), 58% of the explanations that coaches issued occurred during the debrief– either as elaborations of shallow explanations given during problem solving (28%), restatements of problem-solving explanations (6%), or novel justifications of advice given but not explained during problem solving (25%). Further analysis of the entire corpus of debrief dialogues (thirty) revealed that approximately 78% contained discussions which addressed the misconception underlying an error made during problem solving. These discussions were rich in information about how the system under investigation works, and students typically initiated them. Agreement on staging of instruction was 88% (kappa = .85). There was no correlation between type of impasse and staging.

Interventions to address problems with peer interaction. When it is clear that one student knows what the next appropriate action is but progress is still not being made, the coach has a choice. He can simply advise students about what to do, without attending to the obstacle(s) that prevented the student coach from resolving the impasse, or the expert can tailor his advice to fit both levels of diagnosis. As an example of the latter, the expert/coach in Table 4 addresses the solver's error–i.e., a failure to heed a failed diagnostic test–by indirectly confirming the student coach's critique [2, 4]. At the same time, the coach apparently perceives that the student coach is addressing the solver vicariously, and tries to scaffold him in implementing a more direct form of coaching [6, 10]. This intervention thus addresses two problems: one at the student level, and one at the interaction level. The mentor could have chosen to only address the solver's error, by simply telling him that he missed that diagnostic test 2 failed.

How common was it for coaches to address problems at the interaction level? During problem solving, coaches did this during only 25% of one-sided knowledge deficits. However, an additional 25% of interaction-level problems were addressed during the post-practice discussions. Subsequent analysis of the entire corpus of debrief dialogues revealed that there were four main types of interaction-level discussions. The most common (55%) involved explanations by the expert or student coach about why he did not intervene. Typically, this explanation consisted of a description of a mistake that the student coach or expert made while coaching. However, sometimes the student coach or expert kept silent out of amusement–that is, to see how much deeper in trouble the solver would sink! Other types of interaction-level discussions include: the coach's justification for intervening, stated in terms of a problem with peer interaction (20%); the coach's acknowledgement that he knew that students were having trouble, with no explanation as to why he did not intervene (15%); and prompts to students to help their peer more (10%). Agreement on addressing interaction-level problems was 86% (kappa = .72).

Discussion: Implications for CSCL design

Our analysis of the skill of coaching collaboration shows that it is indeed a complex one. Human mentors are challenged to carry out three levels of diagnosis simulateously: at the

task, student and interaction levels. They must also decide how to use their diagnoses to intervene effectively–what to say when, and how to say it. We see three main implications of this study for the design of automated CSCL coaches:

1. *Detecting peer coaching impasses is the easiest task.* As shown in Table 1, voluntary interventions by the coach coincided mainly with cues that do not require deep processing of verbal input: lots of talk but little productive action; little talk and lots of unproductive actions. However, another 20% of voluntary interventions depended on recognizing that a student has not responded to his peer's inappropriate claims. In a CSCL system, this would similarly require recognition of natural language input, or use of structured "natural language" interfaces that allow students to say much of what they want to say in a task domain, while making the system privy to their conversation (Kanselaar & Erkins, 1995). In addition to their interpretive role, these "natural language" interfaces might also prevent certain types of impasses. For example, referential ambiguity and unclear language would probably be reduced. The cost, of course, is slower, restricted communication.

2. *Understanding dialogue participants' verbal interactions can enhance student-level diagnosis, and is crucial for interaction-level diagnosis.* Most student modeling engines in intelligent tutoring systems rely on students' *physical* actions (e.g., in our domain, making measurements, replacing cards) for cognitive diagnosis; they rely little, if at all, on verbal input. Our analysis of cues that enabled us (and coaches) to determine whether an impasse was due to a one-sided, shared, or parallel knowledge deficit shows that most of these cues are linguistic. (See Table 3).

   Although student-level diagnosis can get quite far without the capability to understand verbal input–assuming, of course, the presence of effective inferencing routines–this is not the case at the interaction level. Some amount of text understanding is required–e.g., sentence openers that allow the system to make inferences at the interaction level (e.g., Soller *et al.,* 1999); structured "natural language" interfaces. Although unrestricted, natural language recognition is beyond current technology, our study helps to define the challenge. For example, in order to recognize that the student coach in Table 4 [1] is advising his peer vicariously, the system would need to process his statements at the syntactic, semantic and pragmatic levels. Regarding the semantic level, the system must recognize that "that test" refers to diagnostic test 3. Regarding pragmatics, it must understand that the student's question is more than a request for the expert to interpret the solver's actions (if it is even that); it is predominantly an indirectly expressed critique.

3. *Include a post-practice review session.* As we discussed, the "debrief" dialogues played an important role in instruction at the student and interaction levels. We consider the many discussions of misconceptions underlying faulty actions and interpretations especially interesting, since our research supports prior work on tutoring which showed that, during problem-solving exercises, tutors typically

correct errors rather than diagnose the misconceptions that caused them (McArthur *et al.*, 1990). Determining just how important these discussions are for student learning is a question for future research, as are the questions of how important it is to address interaction-level problems, and when.

Acknowledgements

Bibliography

Gadd, C. S. (1995, July). A theory of the multiple roles of diagnosis in collaborative problem solving discourse. In J. D. Moore and J. F. Lehman (Eds)., *Proceedings of the Seventeenth Annual Conference of the Cognitive Science Society* (pp. 352-357), Pittsburgh, PA.

Hume, G. D., Michael, J., Rovick, A., & Evens, M. (1996). Hinting as a tactic in one-on-one tutoring. *Journal of the Learning Sciences*, *5(1)*, pp. 23-47.

Kanselaar, G., & Erkens, G. (1995). A cooperative system for collaborative problem solving. In Schnase, J.L., and Cunnius, E.L, *Proceedings of CSCL '95: The First International Conference on Computer Support for Collaborative Learning* (pp. 191-194), Bloomington, Indiana. New Jersey: Lawrence Erlbaum Associates.

Katz, S. (1995). Identifying the support needed in computer-*supported* collaborative learning systems. In Schnase, J.L., and Cunnius, E.L, *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., 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.

Leinhardt, G., & Greeno, J. (1986). The cognitive skill of teaching. *Journal of Educational Psychology*, *78*(2), 75-95.

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

Soller, A., Linton, F., Goodman, B., & Lesgold, A. (1999). Toward intelligent analysis and support of collaborative learning interactions. In Lajoie, S.P. and Vivet, M., *Proceedings of Artificial Intelligence in Education 1999* (pp. 75-84), Le Mans, France. Amerstam: IOS Press.

Authors' Addresses

*Sandra Katz (*mailto:katz+@pitt.edu*)*
554 Learning Research and Development Center, University of Pittsburgh; 3939 O'Hara Street; Pittsburgh, PA 15260. Tel. (412) 624-7054. Fax (412) 624-9149.
*Gabriel O'Donnell (*mailto:gabrielo+@pitt.edu*)*
508C Learning Research and Development Center, University of Pittsburgh; 3939 O'Hara Street; Pittsburgh, PA 15260. Tel. (412) 624-7057. Fax (412) 624-9149.