Evaluating Knowledge Construction Dialogs (KCDs) versus minilessons within Andes2 and alone

Stephanie Siler, Carolyn Penstein Rosé, Tim Frost, Kurt Vanlehn
Learning, Research and Development Center
University of Pittsburgh, Pittsburgh, PA 15260
siler@pitt.edu

Peter Koehler
Department of Physics, University of Pittsburgh, Pittsburgh, PA 15260

Abstract
In this paper, we will compare student learning of basic physics concepts when they engage in natural language dialog, specifically, in Knowledge Construction Dialogs (KCDs) with student learning when they simply read about the physics concepts (minilessons). Because KCDs require increased student activity through incremental question-answering, we expected students to learn more from the KCDs than the minilessons. We conducted a study in which we compared student learning from problem-solving in Andes2 with either KCDs or expository minilessons. No learning differences were found. We then conducted a follow-up study in which students either worked through 30 KCDs or read 30 corresponding minilessons, and found higher learning gains for students in the minilesson condition. However, because these results may have been influenced by technical difficulties experienced by students in the KCD condition or student fatigue in the KCD condition, we conducted another study in which we reduced the number of KCDs and minilessons in the intervention phase and fixed the problems leading to server crashes. In this study, there was no difference between conditions in student learning; however, more subjects need to be run before we can make any conclusions regarding the relative effect of student learning from KCDs and minilessons. Finally, we discuss possible explanations for the null results and plans for a future study.

keywords: interactive human-computer dialog

Introduction
When designing intelligent tutoring systems, it is important to apply our understanding of how students learn to its design (e.g., Anderson, Corbett, Koedinger & Pelletier, 1995). One factor that correlates with student learning is student activity. For example, generating rather than reading words has been shown to promote later recall of those words (e.g., Slamecka & Graf, 1978). Promoting students' self-explanations, which includes relating new information to students' prior knowledge, has been shown to promote student learning (e.g., Chi, DeLeeuw, Chiu & Lavancher, 1994; Renkl, 1997). Furthermore, promoting student activity via scaffolding by human tutors, which included incrementally questioning students on a given conceptual topic, was associated with student learning in Chi, Siler, Jeong, Yamauchi and Hausmann (2001). These empirical findings suggest that promoting student activity in a similar manner in intelligent tutoring systems may also increase student learning.

With these principles of student learning in mind, we developed Knowledge Construction Dialogs (KCDs) (Freedman et al., 2000; Rosé et al., 2001). KCDs are interactive directed lines of reasoning that are each designed to lead students to learn as independently as possible one or a small number of concepts. When a question is presented to a student, the student types a response in a text box in natural language. The student may also simply click on Continue, and thus neglect to answer the question. If the student enters a wrong or empty response, the system will engage the student in a remediation sub-dialogue designed to lead the student to the right answer to the corresponding question (refer to the italicized text in the example KCD, below). The system selects a subdialogue based on the content of the student's response, so that incorrect responses that provide evidence of an underlying misconception can be handled differently than responses that simply show ignorance of correct concepts. Once the remediation is complete, the KCD returns to the next question in the directed line of reasoning.
Here is an example KCD with remediation subdialogues italicized:

**Tutor:** Let's think about the difference between speed and velocity. A closely related distinction is that of the difference between distance traveled and displacement from the origin. Take as an example a bumblebee flying from point A to point B by means of a curvy path. If you draw a vector from point A to point B, you will have drawn the bee's displacement vector. What does the displacement vector represent?

**Student:** The distance traveled by the bee from its starting point to its ending point.

**Tutor:** The displacement vector only represents the net change in position accomplished by the bee's motion. So the displacement vector does not tell you how far the bee traveled. It only tells you the net change in the bee's position. If you want to know the actual distance traveled by the bee, what entity do you have to measure the length of?

**Student:** The bee's path.

**Tutor:** Right. The bee's displacement vector is obviously a vector. Is the actual distance traveled by the bee a vector or a scalar?

**Student:** Scalar

**Tutor:** Great. So distance traveled is the scalar counterpart to displacement, which is a vector quantity. Now, can you tell me what is the difference between speed and velocity?

**Student:** Velocity is a vector composed of speed and direction.

**Tutor:** Speed is a scalar. So when you measure the bee's speed, do you want to use the distance traveled, which is a scalar, or the displacement, which is a vector?

**Student:** Distance traveled.

**Tutor:** Good. So the equation for speed is the length of the path traveled by the body divided by duration of the motion, regardless of how curvy the path is, even if the path leads back to where the body started from.

Prior empirical work (Rosé, Jordan, Ringenberg, Siler, VanLehn & Weinstein, 2001) demonstrated that students learned more from the Andes quantitative physics tutoring system (Gertner & VanLehn, 2000) when it was enhanced with KCD capabilities than when it did not include dialog capabilities. However, the KCDs in this experiment had different content than the expository minilessons used in the control condition. In particular, the minilessons only summarized the main point of the corresponding KCD, they did not present the entire line of reasoning employed by the KCD to teach the concept. Thus, the positive effect could have been due to differences in the content rather than the dialogue vs. monologue difference.

To address this possibility, in this paper, we will compare student learning of basic physics concepts from Knowledge Construction Dialogs (KCDs) to student learning from minilessons, non-interactive lessons that contain all of the conceptual content of the corresponding KCDs. Thus, in both conditions, students were exposed to the same lines of reasoning to teach each concept. For each KCD, the corresponding minilesson was systematically derived from the main directed line of reasoning of the KCD, but did not include any of the content from potential remediation subdialogues. The minilessons and the directed lines of reasoning of the corresponding KCDs have about the same average number of words (259 and 264, computed over 15 different minilessons/KCD pairs, a non-significant difference). The content of the KCD directed lines of reasoning and their corresponding minilessons are intended to be virtually identical.

Here is an example minilesson:

Here are a few things to keep in mind when calculating acceleration for a body at rest. Acceleration is change in velocity over time. If velocity is not changing, then there is zero acceleration. Now, if for a finite time interval the velocity remains zero, then it is true that it is not changing. Therefore when a body is at rest during a time interval, its acceleration within that time interval is also zero.

Here is the KCD corresponding to the minilesson:

**Tutor:** Here are a few little things to keep in mind when computing the acceleration vector for a body at rest. Acceleration is change in what over time?

**Student:** velocity.

**Tutor:** Right.

If the velocity is not changing, what is the magnitude of the acceleration?

**Student:** zero.

**Tutor:** Sounds good.
If velocity remains zero within a time interval, how much is it changing?

**Student:** zero.

**Tutor:** Super.

So if a body is at rest during a time interval, what is the magnitude of the acceleration within that interval?

**Student:** 0.

**Tutor:** Correct.

In this paper, we will discuss student learning from KCDs and minilessons both within a computer tutoring system, Andes2, and on their own (outside of any tutoring system). Additionally, we will consider some qualities of the KCDs that are associated with learning gains. Because they are more consistent with the principles of active learning, we expected students to learn more from interacting with KCDs than from reading minilessons.

**Fall 2001 Evaluation**

To determine whether the natural language *interactions* rather than the conceptual content of the KCDs was responsible for student learning gains in the Rosé et al. (2001) study, we ran a study in the Fall of 2001. Subjects for this study were paid volunteers who either were taking or had taken an introductory physics course. Students completed a pretest that consisted of 30 multiple-choice questions designed to assess students' knowledge of the target concepts taught by the KCDs or minilessons that could be elicited during problem-solving and 6 quantitative multi-step problems. The maximum possible score on the multiple-choice section of the pretest was 41. Students were given step-by-step instructions on using Andes2 to solve two warm-up problems, and then were to solve 8 problems in Andes2 on their own. In the intervention phase of this study, we compared two versions of Andes2, Andes2-KCD and Andes2-minilesson. The only difference between these two versions of Andes2 was that a student could receive a KCD in Andes2-KCD following signs the student did not understand a relevant concept, whereas in Andes2-minilesson version the student would instead receive the corresponding minilesson. After the intervention phase, students completed the posttest, which was identical to the pretest.

Of the subjects who took both the pretest and posttest (12 Andes2-minilesson and 14 Andes2-KCDs subjects), the mean pretest score for multiple-choice questions of Andes2-minilesson subjects was higher (but not significantly higher) than for the Andes2-KCD subjects (refer to Table 1, below). The mean posttest score for the multiple-choice questions was marginally significantly higher for Andes2-minilesson subjects (refer to Table 1). No subjects reached the maximum possible score on either the pretest or posttest; thus, there was no ceiling effect. Mean times per problem solved were 0.91 and .84 hours, respectively, a non-significant difference. There was no interaction between condition and either posttest scores or total time on training. There was no difference between conditions after covarying out time on training and pretest scores.

<table>
<thead>
<tr>
<th>Table 1. Mean scores (and standard deviations) for Fall 2001 evaluation (multiple-choice questions)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Andes2-KCD</td>
</tr>
<tr>
<td>Pretest</td>
</tr>
<tr>
<td>Posttest</td>
</tr>
<tr>
<td>t(24) = 1.64, p = .11</td>
</tr>
</tbody>
</table>

To see if a more fine-grained analysis would uncover any differences between conditions, we did the following. For pretest questions that students did not correctly answer, we looked at whether students correctly answered that question on the posttest. Instances in which students received a KCD or minilesson on a topic they did not understand in the pretest were considered learning opportunities. We expected learning opportunities to result in learning more frequently if students had received KCDs or minilessons than if they did not receive such remediation. For the students who completed the posttest, there was a total of 82 learning opportunities for the 14 students in the Andes2-KCD condition (on average, 5.9 per subject) but only 22 learning opportunities for the 9 students in the Andes2-minilesson condition (on average, 2.4 per subject; due to technical difficulties, we lost data of three subjects in this condition). Unfortunately, there weren't enough learning opportunities to compare students' gain scores on individual KCD/minilesson topics (which would control for concept difficulty), so all learning opportunities were aggregated. Students in the Andes2-KCD condition had a 41% likelihood of gaining on a concept given that they received the corresponding KCD,
whereas students in the Andes2-minilesson condition only had a 23% likelihood of gaining, given that they received the corresponding minilesson. This difference in gain, however, was not significant.

This non-significant result may have been due to students in the Andes2-KCD condition not answering all questions in the KCD. Thus, we looked at the log files of the KCD dialogs to see whether subjects who could gain on the KCD topics were actually answering the questions. We excluded the learning events for which students did not answer all questions in the relevant KCD (i.e., there was at least one "I don’t know" or nonsensical response in the log files. "I don’t know" responses in log files indicate that either the subject typed "I don’t know" or simply clicked "Continue", without entering a response, which is recorded in the log files as an "I don’t know" response. A nonsensical response includes non-word responses such as "ll"). We found that some of these students sometimes did not respond to the questions, and, based on the times between turns logged in the KCD dialogs, they did not appear to be reading the questions either. One student often canceled KCDs immediately when they were elicited by typing "cancel". However, it is possible that instances of "I don’t know" or nonsensical responses had different causes, including lack of student effort, neglecting to respond because the answer was known, and actually not knowing the answer. Instances of actually not knowing the answer may lead to cognitive conflict, which is associated with learning (e.g., Dreyfus, Jungwirth, & Eliovitch, 1990; Bearson, Magzamen, & Filardo, 1986; Tudge, 1989). An analysis of the time between turns may help us uncover which of these possibilities is more likely in different cases. Thirty-four learning opportunities were excluded because the student did not attempt to respond to each question in the KCD, leaving 48 learning opportunities for students in the Andes2-KCD condition.

Using only learning opportunities in which students in the KCD condition answered all questions in the KCDs and reached the end of the KCDs, gains were compared between conditions. Reducing our data in this manner may unfairly favor the KCD condition because we have no similar way to reduce the data in the minilesson condition. However, we performed this analysis to determine if lack of student responses may have contributed to the overall null finding. Given that a student in the KCD condition did not know a concept in the pretest and had worked through the entire KCD once in the learning intervention, there was a 41% likelihood of gaining (20 out of 48 learning opportunities). Given that students in the minilesson condition did not know a concept in the pretest but received the minilesson addressing that concept in the learning intervention, they were only 23% likely to gain (5 out of 22 opportunities). Though in the expected direction, again this difference was not significant. The possibility exists that students in the minilesson condition did not always read the minilessons. To address this possibility, we looked at the times minilessons were displayed, which were recorded in the log files. We found a bimodal distribution of times clustered at 4.5 seconds and at 56 seconds. We eliminated the 16 minilesson learning opportunities whose times were clustered around 4.5 seconds. Of the remaining 6 minilesson learning opportunities, students gained on 2 (33%).

It is also possible that students were receiving inappropriate feedback for their responses during KCDs. To test for this possibility, we analyzed roughly one-third of the KCD transcripts and found a 100% precision (the percentage of students’ answers coded as correct that actually were correct) and a 96% recall (the percentage of correct answers identified as correct). Thus, students did receive negative feedback for correct responses 4% of the time. Though infrequent, the impact that receiving negative feedback had on students’ effort and motivation is currently being investigated. Specifically, we are looking at instances in which students received negative feedback on correct responses and seeing whether this was associated with a subsequent increase in students’ “I don’t know” or nonsensical responses. A second type of inappropriate feedback is receiving positive feedback for wrong or off-target responses, which may also confuse students. This is also being investigated.

Additionally, because KCD/minilesson topics were aggregated to achieve statistical power, we looked at the possibility that the concepts KCD students gained on were generally more difficult than the concepts that students in the minilesson condition gained on. Difficulty of a concepts was operationalized as the average probability that the 26 students who completed the posttest demonstrated a lack of understanding of those concepts in the pretest. The average learning opportunity concept was more difficult in the Andes2-minilesson condition than in the Andes2-KCD condition (58% versus 45% chance of answering incorrectly in the pretest, respectively); thus, concept difficulty was not responsible for KCD subjects failing to out-gain minilesson subjects.

Because prior knowledge highly correlates with student learning, a second important factor to consider is differences between conditions in students’ pretest scores. However, the average pretest scores associated with a learning event were similar across Andes2-KCD and Andes2-minilesson conditions (22.4 and 24, respectively).
Spring 2002 Follow-up study

To increase the number of learning opportunities by exposing students to all KCDs and minilessons and to increase the likelihood that students' knowledge gains are prompted by the KCDs and minilessons themselves rather than Andes, we ran a follow-up experiment. We advertised our study on a website that listed experiments available to students taking an introductory psychology class, who were required to participate in experiments for course credit. We also posted flyers around campus. Subjects who were recruited from the psychology subject pool received course credit for their participation; subjects recruited from campus flyers were paid $7/hour. All subjects had taken or were currently taking an introductory physics course. Subjects completed a physics pretest that was similar to the tests taken by subjects in the Fall 2001 evaluation but included additional multiple-choice questions designed to assess students' knowledge of concepts addressed by additional KCDs or minilessons that were not relevant in the Fall 2001 evaluation. Because students were not solving physics problems in this study and thus, we did not expect any learning gain at the problem level, the pretest in this follow-up study also did not include any quantitative problems. After completing the pretest, subjects read through a set of 30 minilessons or worked through the corresponding set of 30 KCDs. Immediately after completing the set of minilessons or KCDs, all students completed the physics posttest, identical to the pretest. Both tests consisted of 44 multiple-choice and short-answer questions designed to tap students' knowledge of the concepts addressed by all of the KCDs and minilessons. The maximum possible score on both the pretest and posttest was 47 (three questions were broken down into 2 sub-parts).

The mean pretest scores of the 11 students in the KCD and the 10 students in the minilesson condition who read through all 30 KCDs or minilessons were similar (refer to Table 2, below). Students in the minilesson condition scored marginally higher on the posttest (refer to Table 2). After covarying students' pretest scores (there was no interaction between students' pretest scores and condition), students in the minilesson condition still out-performed students in the KCD condition, t(18) = 3.8, p < .10.

<table>
<thead>
<tr>
<th>Table 2. Mean scores (and standard deviations)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>Pretest</td>
</tr>
<tr>
<td>Posttest</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

Again, to get a more precise measure of the relative learning benefits specifically from working through KCDs versus reading minilessons, we performed the same fine-grained learning opportunity analysis as in the Fall 2001 Andes2-KCD evaluation. In the KCD condition, students gained on 61 of the 223 opportunities (only 27%), whereas in the minilesson condition, students gained on 84 of the 191 opportunities (44%). Consistent with the overall difference between conditions, this difference in gain was significant, $\chi^2 = 12.5, p < .001$. KCD students gained on 32 of the 118 learning events, or 27%, identical to the rate including all learning opportunities. Once again, students' "I don't know" responses did not appear to be associated with decreased learning.

One explanation for this unexpected result was that the high frequency of server that occurred while students worked through the KCDs interfered with their learning. Another possible factor was that students in the KCD condition, who took longer to finish the 30 KCDs than minilesson subjects took to read, simply became more tired.

Follow-up study to Spring 2002 study

To test these two possibilities, in a second follow-up study, we reduced the number of KCDs and minilessons to 15 and fixed the problems leading to server crashes. Twelve students who were currently taking or had recently taken introductory physics class were recruited from flyers posted around campus. They were paid $7/hour. Twelve subjects were randomly assigned to each condition (6 in each condition). As in the initial follow-up study, students completed the multiple-choice pretest (identical to the pretest given to subjects in the initial Spring 2002 study), worked through the 15 KCDs or read the 15 corresponding minilessons, then immediately completed the multiple-choice posttest (identical to the posttest given to subjects in the initial Spring 2002 study). The maximum possible score on both the pretest and posttest was 47. The mean pretest score of KCD subjects was significantly higher than that of Minilesson subjects (refer to Table 3, below). The mean posttest score of subjects in the KCD was marginally significantly higher than that of the Minilesson subjects (refer to Table 3). The maximum posttest score in the KCD condition was 43; thus, no ceiling effect existed. After covarying pretest scores (again there was no interaction between students' pretest scores and condition), there
was no difference between conditions in posttest scores. Thus, mental fatigue and/or server problems may have contributed to some extent for the initial follow-up results. However, a null finding may simply be due to low number of subjects run in this study.

<table>
<thead>
<tr>
<th>Table 3. Mean scores (and standard deviations)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Pretest</td>
</tr>
<tr>
<td>34.7 (8.8)</td>
</tr>
<tr>
<td>Minilesson condition</td>
</tr>
<tr>
<td>24.3 (3.8)</td>
</tr>
<tr>
<td>t*(10) = 2.63, p &lt; .05</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Posttest</td>
</tr>
<tr>
<td>36.8 (8.2)</td>
</tr>
<tr>
<td>28.5 (5.2)</td>
</tr>
<tr>
<td>t* (10) = 2.11, p &lt; .10</td>
</tr>
</tbody>
</table>

' two-tailed t-tests.

We again performed the same fine-grained analysis as in the prior two studies, in which we looked at subjects' likelihood of gaining on a concept, given they had seen the relevant KCD or minilesson. Because the pretest and posttest both included questions addressing concepts not addressed by the KCDs or minilessons, this analysis was even more sensitive to the effects of minilessons versus KCDs on student learning relative to differences in total test scores. Subjects in the KCD condition gained on 24 of the 51 learning opportunities (or 47%), whereas subjects in the minilesson condition gained on 33 of the 83 learning opportunities (or 40%). Consistent with the total test score results, this difference was not significant. However, before we can conclude no learning differences, we need to run more subjects.

**Qualities of KCDs**

Because generating responses should lead to more effective learning (e.g., Slamecka & Graf, 1978), we looked more closely at how concepts were elicited within the KCDs. For some concepts, students were asked to generate a response for a specific concept. For example, in the KCD that discusses Newton's Third Law, the student is asked to state the relationship between the directions of a force and its reaction force. For other concepts, the student is not questioned on the concept but is told the concept. For example, the centripetal acceleration relationship (\(a = \frac{v^2}{r}\)) is stated near the end of a KCD. For the 15 KCDs given to subjects in the last study, we determined whether the corresponding test questions targeted concepts that students were questioned on or stated in the KCDs. For KCD subjects in both follow-up studies, gains were more likely on questioned than stated concepts (40 out of 82, or 49% of questioned concepts were gained on versus only 13 out of 56, or 23% of stated concepts were gained on, a significant difference, \(z = 9.2, p = .002\)). However, questioned concepts seemed to be less difficult than stated concepts: on average, students in both Spring evaluations were 44% likely to not have known a questioned concept on the pretest but 77% likely not to have known a stated concept.

**Current directions**

In summary, we have no evidence that engaging students in dialog enhanced student learning over having students simply read the directed line of reasoning of the dialog, though it may just be that we didn't run enough subjects to see a difference. However, these results suggest the possibility that the directed lines of reasoning rather than the dialogue led to student learning in the study reported in Rosé, Jordan, Ringenberg, Siler, VanLehn & Weinstein (2001), in which students learned more from the KCD-enhanced Andes than when Andes did not include KCDs. Perhaps responding to questions within the KCDs led students to attend better to the individual steps of the dialog but not to the target concept(s) of the dialogs, which were assessed during testing.

To assess this possibility, we plan to run a study this Fall with the following four conditions: (1) students will read minilessons, (2) students will read minilessons and summarize each one, (3) subjects will work through KCDs, and (4) subjects will work through KCDs and summarize each. We will assess student learning of both target concepts and concepts presented in the individual steps of the KCDs. Our hope is that asking students to summarize after working through each KCD will lead to increases in learning of both types of concepts.

**Acknowledgements**

This research is supported by the Office of Naval Research, Cognitive and Neural Sciences Division MURI Grant N00014-00-1-0600 and NSF Grant 9720359 to CIRCLE, a center for research on intelligent tutoring.
References


