Understanding Expert-Novice Differences in Geometry Problem-Solving Tasks: A Sensor-based Approach

SeungJun Kim  
HCI Institute,  
Carnegie Mellon University.  
5000 Forbes Ave.  
Pittsburgh, PA, 15213  
sjunikim@cs.cmu.edu

Anind K. Dey  
HCI Institute,  
Carnegie Mellon University.  
5000 Forbes Ave.  
Pittsburgh, PA, 15213  
anind@cs.cmu.edu

Vincent Aleven  
HCI Institute,  
Carnegie Mellon University.  
5000 Forbes Ave.  
Pittsburgh, PA, 15213  
aleven@cs.cmu.edu

Abstract
Understanding learner differences with sensors is increasingly important for effective learner modeling. Learner models based on a student’s problem-solving actions and the automated interpretation of those actions have successfully advanced computer tutoring services. However, such transaction level actions provide insufficient detail about higher-rate cognitive variations, which may hold key information about individual differences in cognition and learning, and about factors that differentiate attention-switching strategies and instructional effects between individuals. To fill this gap, we have conducted a user study to investigate causal relationships between learners’ expertise levels and patterns of interaction and attention during learning tasks by using an eye tracker and physiological sensors. In this paper, we validate our experimental test-bed built for inferring learners’ cognitive processing states and diagnosing learning phases with sensors, and present initial results about expert-novice differences revealed in transaction level samples and sensor data streams.

Author Keywords
Learning; Learner modeling; Expertise reversal effect; Intelligent tutoring systems
ACM Classification Keywords
H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

Introduction
Understanding individual differences during in-situ learning is important to support learners effectively (emphasized in [9]). Indeed, effective designs and techniques for low-knowledge individuals can lose their effectiveness and even have negative consequences for more proficient learners (“expertise reversal,” [7]). As interactive tutors should appropriately adapt instructional approaches according to changes in a learner’s ability to acquire cognitive skills [12], comprehensive, real-time learner modeling is important for designing tutoring systems that adapt to the needs of individual learners ([3, 11]).

Although successful approaches to learner modeling were traditionally based on transaction data associated with knowledge components (e.g., Knowledge Tracing [4], Bayesian networks [1], or Knowledge Spaces approaches [5]), recent sensor-based approaches have promised more reactive and accurate learner modeling (e.g., gaze-reactive tutoring [5] or combining gaze data and action data in interactive simulations [8] and in exploratory learning environments [2]).

In this context, we hypothesize that the granularity and accuracy of learner models can be improved by exploring informative aspects of learners’ cognitive states and activities, measured through additional data streams such as those from eye trackers and physiological sensors. Accordingly, we focus on improving our understanding of the correlation of real-time measures of students’ attentional and cognitive load states, extracted from sensor data, with their expertise levels.

Specifically, we introduce a user study conducted to investigate differences in the approach patterns that novice and expert learners use to manage their visual attention. We collected data from 21 novices and 20 experts during geometry problem-solving tasks. Here, we present initial evidence showing transactional and perceptual interactions between geometry expertise and task complexity, and demonstrate that eye-tracking can reveal distinguishable patterns in perceptual and cognitive activities between expert and novice learners at a finer-grained level and it can help identification of quantifiable metrics for future learner modeling.

Experimental method
We now describe our experimental setup to investigate expert-novice differences, revealed in transaction level data and sensor data streams, during computer-mediated geometry problem-solving tasks (Figure 1).

Participants
We recruited 41 participants (age $M=28.4$, $SD=12.09$, age range: 18-65, gender: female 51% and male 49%) split across two learner groups according to their expertise levels: 21 novices (age $M=31.6$, $SD=13.76$, age range: 21-65), who are working or studying in social sciences, design, art or drama and who took geometry in high-school but without getting an A and who have not done any geometry since, and 20 experts (age $M=25.1$, $SD=9.24$, age range: 18-62), who are working in or studying math, engineering, computer science, physics or chemistry and who took geometry in high-school and more than one math course in college.

Test-bed and tasks
Participants were provided descriptions of one geometry concept and three theorems, displayed in a tabbed browsing tool on the left side of the screen, and asked to solve a series of geometry problems displayed on the right side of the test-bed screen (See Figure 2).

Figure 2. Student interface on the right side of the test-bed. allowed participants to write notes around diagrams, select drop-down options and radio buttons. Students’ physical interactions with the interface were logged as transactional data with timestamps.

Each participant was asked to solve seven geometry problems. The first three problems (easy) required only one set of theorems to answer, but the next four problems (difficult) required that all three theorems be used in the correct order (e.g., left-side panel of Figure 2). As the main tasks, participants were asked to 1) identify the correct sequence of solution steps (i.e., measures to compute), 2) provide rationale (select from a list) for their choice, and 3) self-report their confidence at each solution step. The last task assumes that a learner’s confidence at each step represents the variation of cognitive load during each fine-grained step and can be combined for an overall task workload.

After each problem, participants were also asked to self-assess their mental workload and task performance by completing two five-point Likert questions.

Sensor-data collection
During the problem-solving tasks, we measured students’ eye-tracking states and collected psycho-physiological responses by using three sensor devices: 1) a SMI Red250 eye tracker, 2) a BioHarnessTM BT to monitor heart rate, and 3) a LightStone finger sensor to measure GSR (Galvanic Skin Response) (See Figure 1). For the baseline measurement, participants were presented with a 120-second relaxation session before the first problem and 60-second relaxation sessions between problems.

Analysis Result 1 - Transactional interaction between expertise level and task complexity
First, we validated whether the two sets of geometry problems (i.e., easy and difficult) induced distinguishable levels of cognitive load (i.e., high and low) and then whether they actually differentiate task performances between expert and novice learners. For this, we performed a two-way ANOVA analysis by using general linear model multivariate method (question complexity as a within-subject factor and learner group as a between-subject factor) and pairwise comparison tests. The following measures were examined:

Measure 1 & 2 – Error rates
As expected, our problem sets led to more errors in transaction-level outcomes from novice learners as compared to expert learners (See the 1st and 2nd pairs of bars in Figure 3). For example, novices made 2.28
times more errors than did experts in providing correct answers and correct justifications together in identification and reasoning tasks across overall solution steps (1st pair).

Figure 3. Novices (vs. Experts) in the question level analysis (i.e., during an entire question across overall solution steps). Measured values from the expert group were set as 1 (the top-most bar). The numbers in each bar represent the proportions of the measured values from the novice group (N=21) to the values from the expert group (N=20).

In particular, when solving high-complexity questions, expert-novice differences were significant in both task performance measures and post-hoc subjective assessment results (See Figure 3). However, for the low-complexity questions, novices performed similarly to experts when providing reasons for the final measures questioned, though they made 1.8 times as many errors in overall answer accuracy.

Measure 3 – Time taken
On average, participants took 209 seconds (3 minutes and 29 seconds) to answer a geometry problem. In particular, across both learner groups, the amount of time dedicated to high-complexity questions more than quadrupled that dedicated to low-complexity questions (high vs. low = 309s vs. 77s); also, novices spent almost 1.8 times longer than experts (p<0.05), regardless of question complexity (see the 3rd pair of bars in Figure 3).

These measures can provide useful features that can effectively discriminate task workload between experts and novices, and task difficulty between low- and high-complexity questions.

Measure 4, 5 & 6 – Subjective assessment
There were clear expert-novice differences in the subjective assessment results (see last three pairs of bars in Figure 3), with novices having less confidence in their performance than experts. The effect sizes were considerably larger for high-complexity questions (Cohen's d=-1.57 in Q1 and d=-1.80 in Q2).

We assumed that a learner’s overall task workload can be determined from an average of the finer-grained (i.e., step-level) cognitive loads experienced and hypothesized that participants’ self-assessed problem-solving performance (Q1) and task workload (Q2) corresponds with their confidence at each solution step, which may be particularly pronounced when solving high-complexity questions; however, unexpectedly we found that differences in expert-novice average confidence across whole solution steps mostly appeared when solving low-complexity questions rather than high-complexity ones.

In order to track finer-grained variations of cognitive load states associated with the attributes of solution steps (i.e., identifying a measure to use among given measures or an interim measure to focus among a number of unspecified measures), we examined expert-novice differences at solution step levels.
Finer-grained analysis
For this, we examined transaction level logs by defining the interval of a solution step as the time interval from the last transaction on the previous step to the first transaction on the new step. Using the task error rate measure (first two pairs in Figure 4), expert-novice differences were also demonstrated in the solution step level analysis, especially while solving high-complexity questions that demand more complex decision-making and reasoning processes during the tasks. Further, their differences in self-reported confidence at the solution step level were also statistically significant in both low- and high-complexity questions.

The results verified that our test-bed and expert-novice criteria successfully induced distinguishable interactions between expertise level and task complexity, and demonstrated that a finer grained level analysis (e.g., solution step level or better) is needed to explore how to detect and predict variations of a learner's cognitive states in real-time.

Analysis Result 2 - Expert-novice differences in eye-tracking patterns
Next, we investigated relationships between learners’ expertise levels and patterns of visual attention. Specifically, we examined eye-tracking measures related to blinks, fixations, saccades, and scan paths. Given different individual problem-solving times, we focused on normalized metrics such as frequencies and average durations, rather than count or total duration.

Quantitative evidence in the question level analysis
As shown in Figure 5, three of the four measures related to saccades (quick, simultaneous movements of both eyes in the same direction) showed significant expert-novice differences. This implies that learner expertise significantly influences attention switching strategies; further, saccade-related measures may contain derivable and quantifiable features that can demonstrate the relationship between expertise level and in-situ cognitive workload for finer-grained learner modeling (e.g., 2nd derivatives of the measures).

Focus maps and eye-fixation scan-paths
We found that the eye movement patterns between our expert and novice learners differed substantially. As illustrated in Figure 6 (top), experts showed selective eye-fixations in pictorial and textual information and took minimal notes around diagrams, but did not refer to any theorems in the theorem menu (left half of the screen) while solving difficult questions. Nevertheless their eye-fixations were focused on relevant areas, showing few interactions around the diagrams but more intensive eye fixations on important areas of the diagrams at confronted solution steps (Figure 7a and b - top). Conversely, novices navigated most of the theorems regardless of question difficulty (Figure 6 and 7b - bottom) and attended fewer question texts on the top of the diagram during the 1st solution steps (Figure 7a -bottom). For some questions, novices (impossibly) didn't use any of given theorems when providing answers despite frequent eye-fixations on the theorem information (e.g., Figure 7b - bottom). Eye tracking revealed that some novices especially confused interior angles of a circle and the definition of arc measures, and failed to apply the major-minor arc theorem that was necessary in most prior solution steps.

Based on the empirical evidence in eye-tracking measures, we summarized the patterns issued from novices and inferred potential explanations, as follows:

- **Pattern 1**: Incorrect sequence of attention shifts at a fast rate ← Shallow reasoning (need to re-learn)
• **Pattern 2**: Attention wandering between two diagram configurations → Difficulty in reasoning and decision-making (need to know 'where to focus')

• **Pattern 3**: Focusing on the right track, but frequent distraction → Pre-cursor of an impasse

• **Pattern 4**: Consistent focus on the right track with no distraction, but no progress → Impasse state (need an assistance in 'how to compute')

When we examine the focus on relevant vs. irrelevant sub-diagram configuration areas associated with solution steps and resulting ratio metrics (e.g., fixation frequency in the irrelevant areas to the relevant area), the expert-novice differences are statistically significant as well over an entire question and even during a single solution step.

**Limitations and Future Work**

In this work-in-progress paper, we validated our test-bed by demonstrating expert-novice differences in transactional data and eye-tracking measures. Then, we conveyed initial evidence of the feasibility of sensor-based approach for a fine-grained learner modeling.

In future work, we will include full results of expert-novice difference in eye tracking patterns with respect to relevant and irrelevant diagram areas, analysis of physiological measures, and identification of additional features for describing expert-novice differences.

**References**


