Guru: A Computer Tutor That Models Expert Human Tutors

Andrew M. Olney1, Sidney D’Mello2, Natalie Person3, Whitney Cade1, Patrick Hays1, Claire Williams1, Blair Lehman1, and Arthur Graesser1

1 University of Memphis
[aolney,wlcade,dphays,mcwlliams,balehman,a-graesser]@memphis.edu
2 University of Notre Dame
sdmello@nd.edu
3 Rhodes College
person@rhodes.edu

Abstract. We present Guru, an intelligent tutoring system for high school biology that has conversations with students, gestures and points to virtual instructional materials, and presents exercises for extended practice. Guru’s instructional strategies are modeled after expert tutors and focus on brief interactive lectures followed by rounds of scaffolding as well as summarizing, concept mapping, and Cloze tasks. This paper describes the Guru session and presents learning outcomes from an in-school study comparing Guru, human tutoring, and classroom instruction. Results indicated significant learning gains for students in the Guru and human tutoring conditions compared to classroom controls.

Keywords: intelligent tutoring system, expert tutor, biology, conversation.

1 Introduction

Guru is a dialogue-based intelligent tutoring system (ITS) in which an animated tutor agent engages the student in a collaborative conversation that references a multimedia workspace displaying and animating images that are relevant to the conversation. Guru provides short lectures on difficult biology topics, models concepts, and asks probing questions. Guru analyzes typed student responses via natural language understanding techniques and provides formative feedback, tailoring the session to individual students’ knowledge levels. At other points in the session, students produce summaries, complete concept maps, and perform Cloze tasks. To our knowledge, Guru is the first ITS that covers an entire high school biology course.

Guru is distinct from most dialogue-based ITSs, such as AutoTutor [1] or Why-Atlas [2], because it is modeled after 50-hours of expert human tutor observations that reveal markedly different pedagogical strategies from previously observed novice tutors [3]. Our computational models of expert tutoring are multi-scale, from tutorial modes (e.g. scaffolding), to collaborative patterns of dialogue moves (e.g. information-elicitation), to individual moves (e.g. direct instruction) [4]. However, the
importance of tutoring expertise has recently been called into question. In a meta-
analysis, VanLehn [5] examined the effectiveness of step-based ITSs and human tu-
toring compared to no tutoring learning controls matched for content. He reported that
the effect sizes of human tutoring are not as large as Bloom’s two sigma effect [6].
Instead, the effect sizes for human tutoring are much lower \( (d = .79) \), and step-based
systems \( (d = .76) \) are comparable to human tutoring. Even so, the relative influence
of expertise on learning outcomes remains unclear and requires more research.

The present study addresses the effectiveness of Guru in promoting learning gains.
Specifically, how do learning gains obtained from classroom instruction + Guru com-
pare to classroom + human tutoring and classroom instruction alone? We begin with
a sketch of Guru followed by an experiment designed to evaluate the effectiveness of
Guru in an authentic learning context, namely an urban high school in the U.S.

2 Brief Description of Guru

Guru covers 120 biology topics aligned with the Tennessee Biology I Curriculum
Standards, each taking from 15 to 40 minutes to cover. Topics are organized around
concepts, e.g. proteins help cells regulate functions. Guru attempts to get students to
articulate each concept over the course of the session. In this study, a Guru session is
ordered in phases: Preview, Lecture, Summary, Concept Maps I, Scaffolding I, Con-
cept Maps II, Scaffolding II, and Cloze Task. Guru begins with a Preview making the
topic concrete and relevant to the student, e.g. “Proteins do lots of different things in
our bodies. In fact, most of your body is made out of proteins!” Guru’s Lectures have
a 3:1 (Tutor:Student) turn ratio [4, 7] in which the tutor asks concept completion
questions (e.g., Enzymes are a type of what?), verification questions (e.g., Is connec-
tive tissue made up of proteins?), or comprehension gauging questions (e.g., Is this
making sense so far?). At the end of the lectures, students generate Summaries;
summary quality determines the concepts to target in the remainder of the session. For
target concepts, students complete skeleton Concept Maps generated from concept
text [8]. In Scaffolding, Guru uses a Direct Instruction → Prompt → Feedback →
Verification Question → Feedback dialogue cycle to cover target concepts. A Cloze
task requiring students to fill in an ideal summary ends the session.

Guru’s interface (see Figure 1) consists of a multimedia panel, a 3D animated
agent, and a response box. The agent speaks, gestures, and points using motion cap-
ture and animation. Throughout the dialogue, the tutor gestures and points to images
on the multimedia panel most relevant to the discussion, and images are slowly
revealed as the dialogue advances. Student typed input is mapped to a speech act cat-
egory (e.g., Answer, Question, Affirmative, etc.) using regular expressions and a
decision tree learned from a labeled tutoring corpus [9,10]. Guru uses speech act cate-
gory and multiple models of dialogue context to decide what to do next. Thus an af-
firmative in the context of a verification question is interpreted as an Answer, while
an affirmative in the context of a statement like “Are you ready to begin?” is not.
Guru uses a general model of dialogue (e.g., feedback, questions, and motivational
dialogue) and specific models representing the mode of the tutoring session, including