Computers & Education 56 (2011) 574–584

Contents lists available at ScienceDirect

Computers & Education

journal homepage: www.elsevier.com/locate/compedu

Polite web-based intelligent tutors: Can they improve learning in classrooms?

Bruce M. McLaren[a,*], Krista E. DeLeeuw[b], Richard E. Mayer[c]

[a] Human-Computer Interaction Institute, Carnegie Mellon University, Pittsburgh, PA 15213-3891, USA
[b] Knowledge Media Research Center, Tübingen, Germany
[c] University of California, Santa Barbara, USA

A R T I C L E   I N F O

Article history:
Received 9 August 2010
Received in revised form
27 September 2010
Accepted 28 September 2010

Keywords:
Intelligent tutoring systems
Interactive learning environments
Media in education
Pedagogical issues

A B S T R A C T

Should an intelligent software tutor be polite, in an effort to motivate and cajole students to learn, or should it use more direct language? If it should be polite, under what conditions? In a series of studies in different contexts (e.g., lab versus classroom) with a variety of students (e.g., low prior knowledge versus high prior knowledge), the *politeness effect* was investigated in the context of web-based intelligent tutoring systems, software that runs on the Internet and employs artificial intelligence and learning science techniques to help students learn. The goal was to pinpoint the appropriate conditions for having the web-based tutors provide polite feedback and hints (e.g., "Let's convert the units of the first item") versus direct feedback and hints (e.g., "Convert the units of the first item now"). In the study presented in this paper, 132 high school students in a classroom setting, grouped as low and high prior knowledge learners according to a pre-intervention knowledge questionnaire, did not benefit more from polite feedback and hints than direct feedback and hints on either an immediate or delayed posttest, both of which contained near transfer and conceptual test items. Of particular interest and contrary to an earlier lab study, low prior knowledge students did not benefit more from using the polite version of a tutor. On the other hand, a politeness effect was observed for the students who made the most errors during the intervention, a different proxy for low prior knowledge, hinting that even in a classroom setting, politeness may be beneficial for more needy students. This article presents and discusses these results, as well as discussing the politeness effect more generally, its theoretical underpinnings, and future directions.

1. Introduction

The abundance of learning materials and programs available on the web today raises an important question: How can we use this new technology to improve learning? Our research is motivated by the idea that it is important not only to provide easy access to learning technology, as the web surely does, but also to investigate in a scientific manner ways to make that technology more beneficial to learning. The field of intelligent tutoring systems (ITS), computer-based learning systems developed with artificial intelligence techniques (VanLehn, 2006), has been providing tutors on the web for some time now (cf. Aleven, McLaren, & Sewall, 2009; Alpert, Singley, & Fairweather, 1999; Beal, Walles, Arroyo, & Woolf, 2007; Melis et al., 2001), and, at the same time, ITS researchers have used an evidence-based approach to demonstrate impressive improvements in student learning in a range of domains and with different techniques (cf. McLaren, Lim, & Koedinger, 2008; Mostow & Beck, 2007; Rickel & Johnson, 1999; VanLehn et al., 2005). ITS research builds on the long and substantial research on instructional feedback, which has demonstrated both significant learning benefits and failures of feedback (Hattie & Gan, 2011; Kluger & DeNisi, 1996; Shute, 2008). The learning benefits of ITSs have also been traced to specific instructional design principles, such as minimizing cognitive load and using immediate feedback (Koedinger & Corbett, 2006; Shute, 2008). In short, research on intelligent tutoring systems has focused on determining *what* tutors should say to students (i.e., communication content) as well as *when* they should say it (i.e., communication pacing).

In contrast, what about the *way* in which feedback is presented to students and, as a consequence, how (and whether) students perceive software tutors as learning partners? Much less research has been done on how best to incorporate *social cues*, such as polite wording, which

* Corresponding author. Tel.: +1 412 268 8278; fax: +1 412 268 1266.
E-mail addresses: bmclaren@cs.cmu.edu (B.M. McLaren), k.deleeuw@iwmkmrc.de (K.E. DeLeeuw), mayer@psych.ucsb.edu (R.E. Mayer).

0360-1315/$ – see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.compedu.2010.09.019

may be an essential element in student–tutor interactions. Our hypothesis is that intelligent tutors should not only exhibit *cognitive intelligence*—knowing what to say and when to say it—but also *social intelligence*—knowing how to say it. The *politeness principle* – the idea that people learn more deeply when instructional support is presented in polite style – is the particular focus of our investigation into social intelligence. A basis for the politeness principle is suggested, first, by the politeness theory of Brown and Levinson (1987). In their studies of politeness across cultures, they identified the key attributes of *positive face*, liking or admiring of another and/or cooperative and suggestive of a common goal (i.e., supporting another's self-esteem), and *negative face*, being respectful of another's right to make his or her own decisions (supporting another's freedom to act). A second underpinning of our work is the media equation theory of Reeves and Nass (1996) – the idea that people "respond socially and naturally to media" (Reeves & Nass, 1996, p. 1), thereby reacting to and treating a computer as a real person. Finally, the social agency theory of Mayer et al. (Mayer, 2005a, 2009) suggests that instructional messages – including feedback and hints from web-based tutors – should be presented in a way that provides social cues, such as polite wording. In this way, the learner may accept the tutor as a conversational partner, resulting in increased effort to learn the given material, which in turn leads to a better learning outcome.

In the study reported in this paper, we had three goals in testing the politeness principle. First, we sought to carefully and empirically test the principle, following from the approach and findings of prior studies, some of which demonstrated clear learning effects (McLaren, DeLeeuw, & Mayer, in press; Wang et al., 2008), some which demonstrated weaker effects (Wang & Johnson, 2008) and some which did not show any effects (McLaren, Lim, Yaron, & Koedinger, 2007). Second, we intended to test the specific conditions that influence the effectiveness of the principle. In particular, we were interested in whether the effect could survive transition from a lab environment, where it previously had been successfully employed to promote learning (McLaren et al., in press; Wang et al., 2008; Wang & Johnson, 2008), to a classroom situation where it had not been successful previously (McLaren et al., 2007). We also intended to test whether low prior knowledge learners benefit more from polite feedback than high prior knowledge learners, as they had in an earlier study we conducted (McLaren et al., in press). Finally, a key goal was to test the politeness principle applied in the context of the worldwide web, rather than within a standard computer software environment. Because the web is omnipresent in today's classrooms, determining whether, and under what conditions, politeness makes a difference to learning with web-based materials could open the door to much wider, and scientifically verified, use of the politeness principle within intelligent tutoring systems.

In the study described in this paper, 132 U.S. high school students who were enrolled in a chemistry class learned from a web-based intelligent tutor designed to teach stoichiometry, a subtopic of chemistry. Following the learning phase, students completed an immediate posttest and a delayed posttest one week later, with both near transfer and conceptual questions on both tests. Some students worked with a version of the Stoichiometry Tutor that provided polite problem statements, feedback, and hints, while others worked with a version of the tutor that provided direct problem statements, feedback, and hints. The polite versions of problem statements, feedback, and hints are based on the face-saving techniques of Brown and Levinson (1987), described briefly above and in more detail in the next section. The study took 3 to 5 hours for students to complete, with most of the work occurring in the classroom (The only work that occurred outside of the classroom was by students who missed class.). The software tutor was developed using authoring software specifically designed to build web-based intelligent tutors (Aleven, McLaren, Sewall, & Koedinger, 2009).

In this paper, we briefly review the theoretical underpinnings of the politeness principle, describe in more detail the empirical studies that have been done to date with the principle, describe the current study and its results, and discuss what we have learned about the politeness principle thus far.

2. How theory informs our operationalization of the politeness principle

According to Brown and Levinson (1987), politeness has a role in all human culture. *Face* is the public presentation of such politeness. Positive face is "the want of every member that his wants be desirable to at least some others," (Brown & Levinson, 1987, p. 62) or, in other words, the desire of people to be appreciated and approved of by others. Positive face is also characterized by the desire to work cooperatively with others. Negative face, on the other hand, is "the want of every 'competent adult member' that his wants be unimpeded by others," (Brown & Levinson, 1987, p. 62), that is, a person's freedom of autonomy. Positive face refers to a person's self esteem, while negative face refers to a person's freedom to act. According to Brown and Levinson, these two components of human desire are fundamental to any social interaction, and politeness is an attempt by the participants in those interactions to maintain each other's faces. Brown and Levinson have documented the similar ways in which people from diverse cultures use politeness tactics for making requests that minimize threats to both positive and negative face.

Our goal in the present research was to phrase the problem statements, hints, and feedback of our intelligent tutor so as not to threaten the positive or negative face of the student user, in a manner suggested by the work of Brown and Levinson. For example, we worded hints of our polite tutor to reduce the threat to positive face by presenting cooperative statements (e.g., "Let's calculate the result now" or "Let's convert the units of the first item") and reduce threats to negative face by allowing freedom of action (e.g., "Do you want to put 1 mol in the numerator?" or "You may want to convert the units of the first term"). Contrast this with the direct version of our tutor that uses more direct wording that threatens positive face by being less cooperative (e.g., "The goal here is to calculate the result") and threatens negative face through authoritative statements (such as "Put 1 mol in the numerator"). In other words, the theoretical motivation for using polite tutors in the present study was to prime the learner for social cooperation which will lead, we hypothesize, to deeper learning.

A second theoretical foundation of our work is the media equation theory of Reeves and Nass (1996), which proposes that people can be induced to relate to a computer as if it is a real person. When social cues are present, Reeves and Nass claim, people easily accept a computer as a social partner, a claim that is supported by their study in which people learning with computers treated the computers politely, similar to the way human beings would be treated. Both Reeves and Nass (1996) and Nass and Brave (2005) provide evidence that people need only a minimal amount of priming to accept a computer as a social partner. Grice (1975) argues that the speaker in a conversation agrees to generate a message that is intended to make sense to the listener (i.e., the speaker agrees to be clear, relevant, concise, and truthful), and the listener agrees to exert effort to make sense of the message. When a learner accepts a computer tutor as a social partner, the learner views a tutor's message as part of a conversation, which is subject to the rules of conversation—including a commitment by the learner to try to make sense of what the tutor is saying. If the learner has made that commitment, he or she should also process the information more deeply, leading to better learning. In the current study, we seek to use polite feedback and hints as a way to encourage students to view a web-based tutor as a social partner and to prime the conversational stance.

Finally, Mayer et al. (Mayer, 2005a, 2009) have proposed social agency theory as an extension of the cognitive theory of multimedia learning. Social agency theory is based on the idea that instructional messages – including feedback and hints from web-based tutor – may be presented in a way that does or does not involve social cues (e.g., does or does not use polite conversational style). When a tutor's message contains appropriate social cues, such as polite wording, the learner accepts the tutor as a conversational partner, which results in increased effort to engage in cognitive processing aimed at making sense of the tutor's message, thereby creating a higher quality learning outcome. When the tutor's message does not contain social cues, such as when the message is directly or authoritatively expressed, the learner is less likely to accept the tutor as a conversational partner, and therefore the learner is less likely to work hard to make sense of the tutor's message, resulting in a lower quality learning outcome. The cognitive processes that lead to better learning are spelled out in the cognitive theory of multimedia learning (Mayer, 2005a, 2009), and include selecting relevant information, mentally organizing it into a coherent structure, and integrating it with other knowledge.

Based on the sum of these theories, we hypothesize that students who work with polite tutors will learn more, both about the specific material presented and the concepts underlying that material, than will students who learn with direct tutors. We also hypothesize that this politeness effect will be strongest for students who have low rather than high prior knowledge. Students with high prior knowledge are more likely to engage in deep cognitive processing during learning without social inducement, and, in fact, polite inducement could even be distracting to these students. However, students with low prior knowledge are more likely to need some inducement, such as being drawn into a social interaction with their tutor, to engage in deeper processing.

3. Relevant previous work with intelligent tutors and politeness

Politeness has been an area of interest within the field of intelligent tutoring systems since at least the mid 1990s. The earliest work may be that of Person, Kreuz, Zwaan, and Graesser (1995) in which they found evidence that politeness strategies are commonly used in one-on-one tutoring interactions between humans, although not always effectively. Their study of human tutoring dialogues suggests that politeness could, under some circumstances and in different domains, inhibit effective tutoring. They also found that different steps in the tutoring process appear to be more or less likely to benefit from politeness. However, these early findings were observational, not subjected to large-scale empirical study, and also not tested with software tutors.

More recently, Mayer, Wang, Johnson and colleagues have performed a series of studies to investigate whether politeness in educational software, in the form of positive and negative face-saving feedback, can better support learners (Mayer, Johnson, Shaw, & Sandhu, 2006; Wang et al., 2008). They implemented positive and negative face feedback in the context of the Virtual Factory Teaching System (VFTS), a factory modeling and simulation tutor. In a *polite* version of VFTS they have developed, constructions such as, "You could press the ENTER key" and "Let's click the ENTER button" were used. Such statements are arguably good for positive face, as they are likely to be perceived as cooperative and suggestive of a common goal, as well as for negative face, as they are also likely to be perceived as respectful of the student's right to make his or her own decisions. In the *direct* version of VFTS, the tutor used more imperative, direct feedback such as, "Press the ENTER key" and "The system is asking you to click the ENTER button." These statements are arguably not supportive of positive face, as they do not suggest cooperation, or of negative face, as they are likely to be perceived as limiting the student's freedom.

In a preliminary study (Mayer et al., 2006) students were asked to evaluate the threat to negative and positive face of a tutor's statements. The results indicated that learners are very sensitive to politeness in tutorial feedback, and that learners with less computer experience react to the level of politeness in language more than experienced computer users. In the follow-up study run by Wang et al. (2008) in which 37 students were randomly assigned either to a polite tutor group or to a direct tutor group, students who used the polite tutor scored significantly higher on a posttest. Importantly, the politeness effect was obtained for non-engineering students but not for engineering students, thus pointing to the notion that students with less prior knowledge of a domain are more susceptible to positive learning effects from the politeness principle. Wang and Johnson (2008) also observed a significant learning effect due to politeness with a foreign language tutoring system for a particular type of question, utterance formation questions, in which participants answer a question by recording their own speech. The participants in this study were paid volunteers, largely without significant prior language training, and thus, as in the Wang et al. (2008) study, low prior knowledge learners. In our own earlier lab study with university students (predominantly psychology majors) using the Stoichiometry Tutor (McLaren et al., in press), there was a significant pattern in which students with low prior knowledge of chemistry performed better on subsequent problem-solving tests if they learned from the polite tutor rather than the direct tutor ($d = 0.64$ on an immediate test, $d = 0.50$ on a delayed test), whereas students with high prior knowledge showed the reverse trend ($d = -0.58$ for an immediate test; $d = -0.21$ for a delayed test). In sum, these studies suggest, first, that the level of politeness in the system feedback of a tutoring system can make a difference in motivating students and promoting better learning and, second, that the effect tends to be stronger, and perhaps only useful for students who have less knowledge in the particular domain of interest.

Yet not all research supports the idea that politeness in intelligent tutoring systems will benefit learning. For instance, the Wang and Johnson (2008) study discussed above, while obtaining a politeness effect for utterance formation questions, did not uncover a politeness learning effect on subjects' overall score, or an effect on multiple-choice questions or questions involving matching of phrases in Iraqi Arabic (the language being tutored) to translations in English. In an earlier classroom study involving the Stoichiometry Tutor McLaren et al. (2007), as opposed to the lab study of McLaren et al. (in press), did not find a politeness effect for high school students in a classroom setting. Although the polite group performed slightly better than the direct group on a posttest, the difference was not statistically significant. Why did the earlier experiment not obtain a politeness effect whereas both our own, as well as other experiments did? One potentially important difference is that the learners in this experiment were students taking a college prep chemistry course with a strong and recent background in chemistry, whereas the learners who produced a politeness effect in the previous experiments were largely unfamiliar with the material. Another key difference is the setting: The McLaren et al. (2007) study was conducted in classrooms, whereas all of the other politeness studies, including our own (McLaren et al., in press), were lab-based studies. In the present experiment, we further explore both the use of the Stoichiometry Tutor in a classroom setting and the idea that low prior knowledge students are most likely to display a politeness effect.

4. Method

4.1. Design

We conducted a study with a 2x2 between-subjects factorial design. One factor was politeness, with one level *polite* instruction (i.e., use of the Polite Stoichiometry Tutor) and the other level *direct* instruction (i.e., use of the Direct Stoichiometry Tutor). The other factor was modality of feedback, with one level being *text only* and the other level *audio-and-text*.[1]

4.2. Participants and conditions

One hundred and thirty-two (132) high school students (72 female and 60 male) in five chemistry classes in three suburban high schools in two states (Massachusetts and New Jersey) participated. There were sixteen additional students who at least partially participated, but scored 0 or very nearly 0 on one or both of the two posttests; these students were eliminated from consideration. We also eliminated two students because the number of total messages they saw, hints requested plus error messages, was more than 3 standard deviations above the mean, indicating they were "gaming" the system (i.e., trying to simply get the answers in the last hint, see Baker et al., 2008; Wood & Wood, 1999). The study materials were used as a replacement for normal lectures and class work on stoichiometry within the five high school classes, and the three participating teachers were given and used the immediate and delayed posttests as class grades for their students.

Students were randomly assigned to one of the four conditions of the 2 × 2 design (polite/text, polite/audio, direct/text, and direct/audio). The number of students that were assigned to each condition, and in total to the polite and direction conditions, is shown in Table 1.

Table 1
Distribution of subjects across conditions.

	Polite Instruction	Direct Instruction
Text Only	Polite/text (33)	Direct/text (32)
Audio and Text	Polite/audio (31)	Direct/audio (36)
	Polite (64)	Direct (68)

4.3. Materials and procedure

Table 2 provides an outline of the materials and procedure used in the study. The left-hand column indicates the materials presented to and completed by students in the polite conditions (i.e., both polite/text and polite/audio) and the right-hand column indicates the materials presented to and completed by students in the direct conditions (both direct/text and direct/audio). The rows between the thick horizontal lines in the middle of the table represent the *intervention materials*; these varied by condition, as further indicated by the highlighting of the direct materials on the right. All materials were completed online, within a web browser, in the order shown in Table 2. Students used school-provided computers and headphones, so students could privately listen to the videos and hear audio feedback and hints. All participants were given user-IDs and passwords that allowed them to logoff and log back on whenever desired.

Because of the usual difficulties in using and tightly controlling classroom time, the study materials of Table 2 were tackled mostly, but not exclusively, during teacher-monitored classroom time. In a few cases, due to absences or insufficient classroom time, the consent form, questionnaires, videos, and intervention materials were completed (or viewed) outside of regular classroom time, at school or home. These materials took students between 60 and 120 min to complete. Two posttests were administered, one immediate and one delayed by one week. Each posttest took between 45 and 60 min to complete. All of the students took the posttests in class.

The pre-questionnaire contained basic demographic questions, as well as questions about the student's background in and understanding of chemistry. The chemistry questions are shown in Table 3. Answers to these questions were used to separate students into low and high prior knowledge groups in subsequent analyses. For the first question a score of 1 ("Far below average") to 5 ("Highly above average") was given to each student. For the second question a score between 0 and 10 was given to each student, based on whether they selected "None of the above are true" (0) or the number of items selected, between 1 and 10. The scores of the two questions were added together and the mean of all the students' scores was calculated. All students who scored below the calculated mean of 11.3 were classified as "low prior knowledge learners," while all students above the mean were classified as "high prior knowledge learners." Note that we did not administer a pretest due to the possibility of "testing effects" (Johnson & Mayer, 2009; Roediger & Karpicke, 2006), the well-studied phenomenon in which student performance improves from taking a test. Since this raises the possibility of washing out learning effects due to the intervention, we elected to instead ask students to self assess their knowledge.

After completing the pre-questionnaire, the students worked on the intervention materials, including videos and the Stoichiometry Tutor, that were specific to their condition, as shown in Table 2. The videos were interspersed throughout the materials and were short (1–4 min), presenting various background materials on chemistry concepts relevant to stoichiometry (e.g., molecular weight, dimensional analysis), as well as tips on how to solve stoichiometry problems (e.g., problem solving strategy). As indicated in Table 2, the language used in the videos (i.e., the narration) was specific to condition – polite language was used in the polite condition videos (e.g., "Let's discuss how

[1] While we originally intended to investigate the learning benefit of providing feedback and hints with a human voice and printed text versus text alone, we later decided that our design had a redundancy effect (Mayer, 2005b) that would make positive learning effects unlikely. Thus, we focused only on politeness in this study and only the polite vs. direct aspect of the study will be discussed in the remainder of this paper.

Table 2
Materials used and design of the study.

Polite	Direct
Web-based Consent Form	Web-based Consent Form
Pre-Questionnaire	Pre-Questionnaire
Five videos, in ***polite*** language: • Introduction to the Stoichiometry Study • Overview of the Tutor and Interface • Stoichiometry Problem Solving Strategy • Dimensional Analysis • Significant Figures	Five videos, in ***direct*** language: • Introduction to the Stoichiometry Study • Overview of the Tutor and Interface • Stoichiometry Problem Solving Strategy • Dimensional Analysis • Significant Figures
Polite Stoichiometry Tutor - Problem # 1	Direct Stoichiometry Tutor - Problem # 1
Polite Stoichiometry Tutor - Problem # 2	Direct Stoichiometry Tutor - Problem # 2
Video: Molecular Weight (in ***polite*** language)	Video: Molecular Weight (in ***direct*** language)
Polite Stoichiometry Tutor - Problem # 3	Direct Stoichiometry Tutor - Problem # 3
Polite Stoichiometry Tutor - Problem # 4	Direct Stoichiometry Tutor - Problem # 4
Video: Composition Stoichiometry (in ***polite*** language)	Video: Composition Stoichiometry (in ***direct*** language)
Polite Stoichiometry Tutor - Problem # 5	Direct Stoichiometry Tutor - Problem # 5
Polite Stoichiometry Tutor - Problem # 6	Direct Stoichiometry Tutor - Problem # 6
Video: Molarity (in ***polite*** language)	Video: Molarity (in ***direct*** language)
Polite Stoichiometry Tutor - Problem # 7	Direct Stoichiometry Tutor - Problem # 7
Polite Stoichiometry Tutor - Problem # 8	Direct Stoichiometry Tutor - Problem # 8
Polite Stoichiometry Tutor - Problem # 9	Direct Stoichiometry Tutor - Problem # 9
Polite Stoichiometry Tutor - Problem # 10	Direct Stoichiometry Tutor - Problem # 10
Post-Questionnaire	Post-Questionnaire
Video: Introduction to Post Test	Video: Introduction to Post Test
Immediate Posttest: 8 Problems (4 near transfer; 4 conceptual)	Immediate Posttest: 8 Problems (4 near transfer; 4 conceptual)
Delayed Posttest: (One week later) 8 Problems (4 near transfer; 4 conceptual)	Delayed Posttest: (One week later) 8 Problems (4 near transfer; 4 conceptual)

composition stoichiometry allows us to determine the make-up of a molecule …") and direct language was used in the direct condition videos (e.g., "Composition stoichiometry refers to the make-up of a molecule…").

Fig. 1 shows the Stoichiometry Tutor, developed using the Cognitive Tutor Authoring Tools (Aleven, McLaren, & Sewall, 2009; Aleven et al., 2009), as well as an example of both its polite (Fig. 1a) and direct (Fig. 1b) version. Solving a stoichiometry problem involves understanding basic chemistry concepts, such as unit conversions (e.g., 1 $g = 1000$ mg, as in Fig. 1) and the mole, and applying those concepts in solving simple algebraic chemistry equations. To solve problems, the student must fill in the terms of an equation, cancel numerators and denominators appropriately, self-explain the reason for each term of the equation (e.g., see the entry of "Given Value" below the first term of Fig. 1a and b), and calculate and fill in a final result. The student can request hints by selecting the "Hint" button in the upper right-hand corner of the interface. If the number typed (or unit or substance or reason selected) is correct, the typed (or selected) information appears in

Table 3
Chemistry knowledge questions on the pre-questionnaire used to divide students between low and high prior knowledge learners for subsequent analyses.

Question	Possible Answers
Please rate your overall knowledge of chemistry	• Highly above average • Above average • Average • Below average • Far below average
Please indicate the items that apply to you	• I plan to major in chemistry • I know what the 2 stands for in H_2O • I know what a mol is • I have heard of Avogadro's number • I know what Na stands for • I know what mL stands for • I know how many significant figures are in 0.0310 • I know how many grams are in a kg • I know what stoichiometry is • I know the difference between an atom and a molecule • None of the above are true

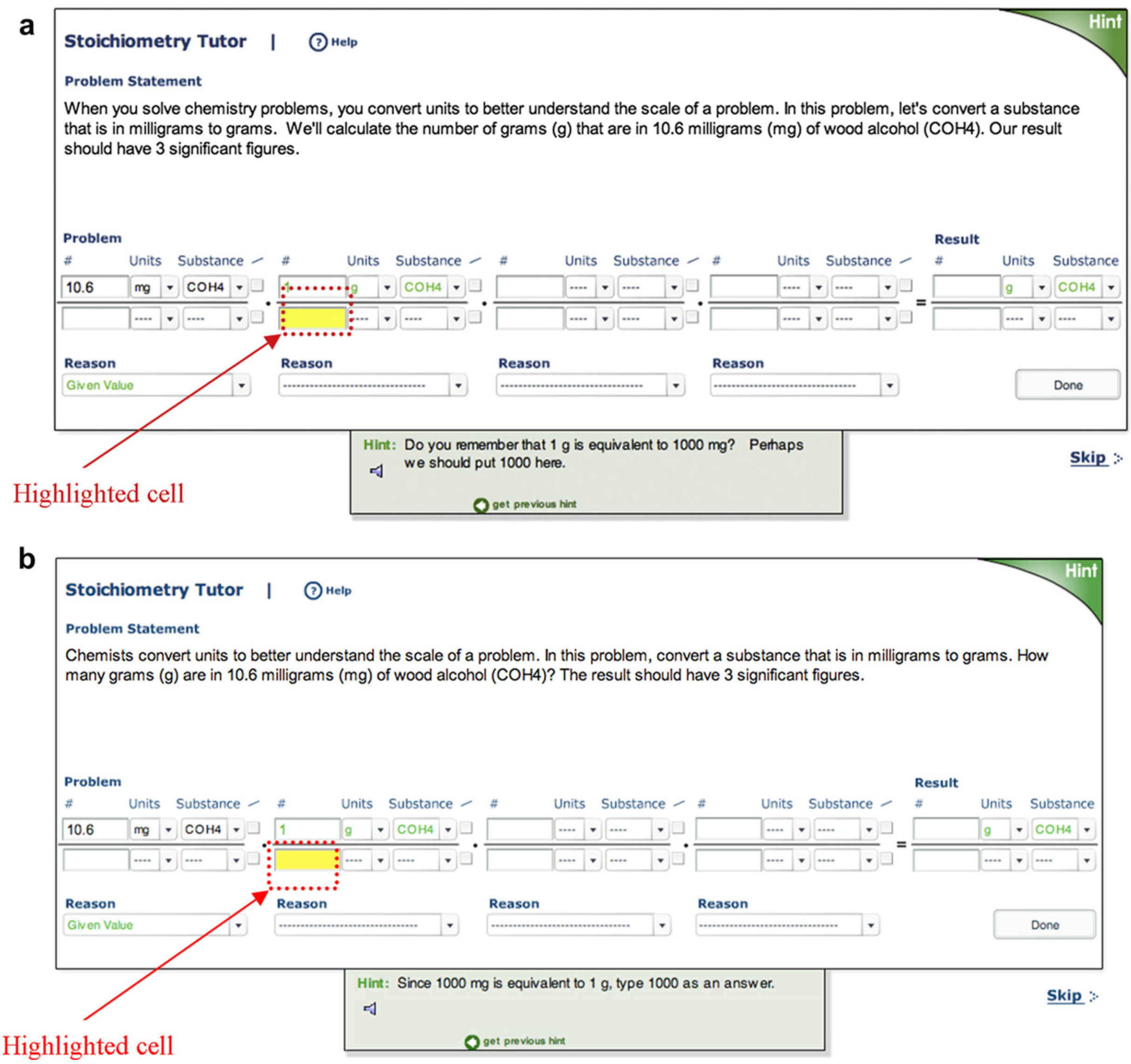

Fig. 1. (a) The Polite Stoichiometry Tutor and an example of a polite hint (b) The Direct Stoichiometry Tutor and an example of a direct hint.

Table 4
Examples of language differences between the polite and direct versions of the Stoichiometry Tutor.

	Polite Stoichiometry Tutor	Direct Stoichiometry Tutor
Problem Statements	Can we calculate the number of grams of iron (Fe) that are present in a gram of hematite (Fe_2O_3)? Our result should have 5 significant figures.	How many grams of iron (Fe) are present in a gram of hematite (Fe_2O_3)? The result should have 5 significant figures.
Hints	Let's calculate the result now.	The goal here is to calculate the result.
	Do you want to put 1 mol in the numerator?	Put 1 mol in the numerator.
	Let's convert the units of the first item	Convert the units of the first item now.
	Shall we calculate the result now?	The tutor wants you to calculate the result now.
Error Messages	You could work on a composition stoichiometric relationship in this term. Are grams part of this relationship?	This problem involves a composition stoichiometric relationship in this term. Grams are not part of this relationship.
	Won't we need these units in the solution? Let's not cancel them, Ok?	No, these units are part of the solution and should not be cancelled.
	Are you sure Molecular Weight is part of this problem? Maybe there is another reason for this term?	No. Molecular Weight is not part of this problem. Select another reason for this term.
Success Messages	Super job, keep it up.	[None]

a green font. If it is incorrect, it appears in red. Occasionally, the polite tutor gives success feedback, following Brown and Levinson's notion of "positive politeness" through approval (Brown & Levinson, 1987, p. 102, Fig. 3).

In both Fig. 1a and b, the student has requested a hint after they have correctly solved several steps of the given problem. The hint, which appears at the bottom of the interface, refers to the highlighted cell in the interface. The hints the tutor gives provides progressively more information for solving the problem, with the last hint on each step providing the final answer for that step (a "bottom-out hint", Fig. 1a and b are examples of such hints). The tutor also provides context-specific error messages when the student makes a mistake during problem solving. An earlier paper (McLaren et al., 2007) provides further detail about the tutor. Table 4 provides examples of problem statements, hints, error messages, and success messages given by the Polite and Direct Stoichiometry Tutors from a corpus of over 4000 messages.

After completing all of the intervention videos and problems, the participants responded to a web-based questionnaire that asked about the effectiveness, helpfulness, and the usability of the tutor and then completed the first test (the immediate posttest). This posttest contained eight problems, four of which were of the same type and had the same user interface as the practice problems (but without hints or error feedback; these are *near transfer* problems) and four of which were more conceptual questions for which participants provided a final result in one or two boxes (*conceptual questions*). Fig. 2a shows an example of a near transfer problem and Fig. 2b shows an example of a conceptual question. Approximately one week later, participants completed a second (delayed) posttest in class. This test also contained eight problems – four near transfer questions, four conceptual questions – that were isomorphic to, but different from the immediate test. The order of the two tests was counterbalanced across participants (i.e., one-half of the participants received test A as the immediate test and test B as the delayed test and vice versa).

Both the immediate and delayed posttests were scored by calculating an average per problem (i.e., dividing the number of correct steps the student took on a single problem by the total number of steps for that problem). Separate scores were also calculated for the near transfer and conceptual portions of the posttests.

5. Results

Table 5 shows the means and standard deviations for the main dependent variables for low and high prior knowledge students who used either the polite or direct tutor. It also shows the Cohen's d values, comparing the polite and direct conditions for low and high prior knowledge groups and with respect to the specific dependent variables. The immediate and delayed test scores were calculated using the eight posttest problems, as described in the previous section. *Hints* refers to the average number of hints requested and seen by students during the intervention in each condition, *Errors Seen* refers to the average number of error messages seen by students during the intervention in each condition,[2] and *Hints + Errors Seen* is a combination of these two values. Note, first, that none of the Cohen's d values rises above the level of a small effect (Cohen, 1988).

To compare the polite and direct groups in more statistical depth and with respect to the two posttest scores (i.e., Immediate and Delayed), we performed an analysis of variance, with Hints + Errors Seen as a covariate to control for the amount of exposure to the tutor, since students who saw more hints and errors also had more exposure to the intervention. Although students who used the Polite Stoichiometry Tutor performed at least slightly better than those who used the Direct Stoichiometry Tutor as compared within each of the prior knowledge groups per test (see Table 5), the ANCOVA revealed no statistically significant differences between any of the polite and direct groups nor any interactions with knowledge level.

To compare the polite and direct groups with respect to Hints, Errors Seen, and Hints + Errors Seen, we performed an analysis of variance with no covariate, since using Hints + Errors Seen as the covariate violates statistical assumptions (i.e., the dependent measures would then overlap with the covariate). These tests also revealed no significant differences between the polite and direct groups and no significant interactions with knowledge level. In summary, there were no significant differences between the polite and direct groups, divided into low and high prior knowledge learners, with respect to any of the dependent variables of interest.

As an alternative analysis, we considered the number of errors made during the intervention as a proxy for prior knowledge (i.e., higher prior knowledge students are less likely to make errors and ask for hints than are low prior knowledge students). We divided students into three "error" groups based on the total number of errors they made during the intervention and compared polite and direct conditions within each group with respect to our dependent variables. The low error group consisted of those students who made 81 or fewer errors ($n = 45$: Direct: 29; Polite: 16), the medium error group made between 82 and 132 errors ($n = 43$: Direct: 24; Polite: 19), and the high error

[2] Note that there is a difference between error messages seen and errors actually made. More specifically, not all errors made by students lead to the tutor displaying a text message; in many cases an error is simply indicated by the answer turning red. Only more common errors elicit messages by the tutor.

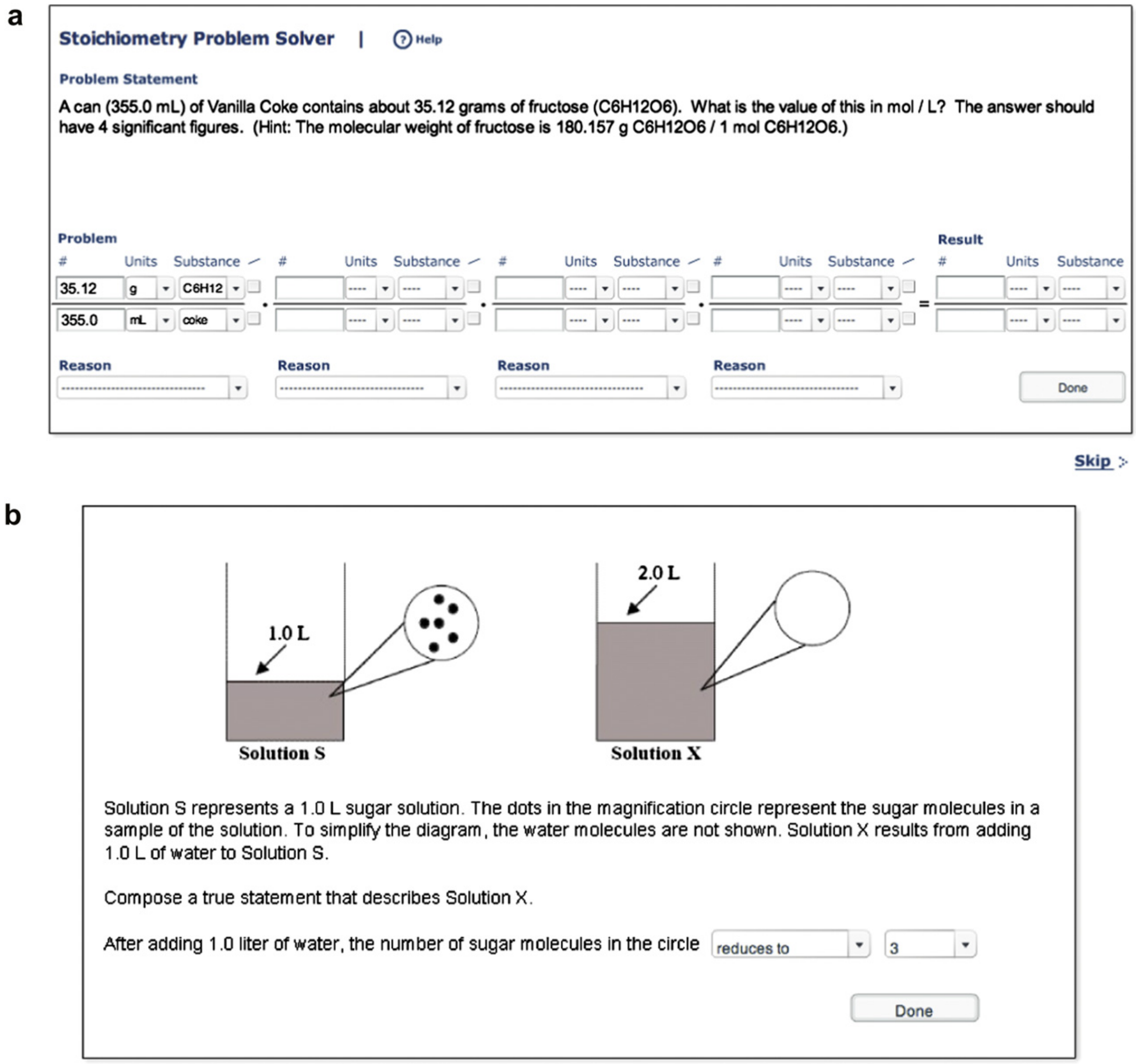

Fig. 2. (a) Example of a near transfer posttest problem. Notice that, unlike Fig. 1a and b, there is no "Hint" button in the upper right-hand corner, meaning that students are on their own in solving this test problem (b) Example of a conceptual posttest problem.

group made 133 or more errors ($n = 44$; Direct: 15; Polite: 29). Table 6 shows the descriptive results of these analyses. Note that the Cohen's d values are all small, except for the one comparing the polite and direct conditions for the immediate posttest within the high error group; this d value of 0.66 would be considered in the medium to medium-high range (Cohen, 1988).

To compare the polite and direct groups in more statistical depth, we conducted a 2 (polite vs. direct) × 2 (text vs. audio-text) × 3 (low, medium, and high error rates) ANOVA for each of our independent variables. For the immediate test, we found no significant main effect of politeness and no interaction of politeness with error rate. However, because we hypothesized that students who make more errors (i.e., lower knowledge students) would be more affected by the politeness manipulation, we also investigated the simple main effects within each Error group by conducting paired comparisons, which compared the polite group to the direct group within each Error group. This

Table 5
Comparison of the polite and direct conditions, divided into low and high prior knowledge groups, with respect to various dependent variables.

	Low Prior Knowledge			High Prior Knowledge		
	Polite Mean (sd)	Direct Mean (sd)	Cohen's d	Polite Mean (sd)	Direct Mean (sd)	Cohen's d
Immediate Posttest	0.61 (0.21)	0.56 (0.19)	0.24	0.67 (0.18)	0.65 (0.23)	0.12
Delayed Posttest	0.60 (0.18)	0.57 (0.20)	0.17	0.69 (0.16)	0.67 (0.15)	0.13
Hints	51.03 (45.74)	59.34 (62.67)	−0.15	31.61 (39.28)	24.19 (25.05)	0.23
Errors Seen	84.88 (43.04)	76.06 (42.77)	0.21	77.55 (51.48)	62.06 (30.64)	0.38
Hints + Errors Seen	135.91 (77.18)	135.41 (88.89)	0.01	109.16 (73.82)	86.25 (41.86)	0.40

*None of the comparisons between polite and direct means resulted in a significant difference, i.e., $p < 0.05$.

Table 6
Comparison of the polite and direct conditions, divided into low, medium, and high error groups, with respect to immediate and delayed posttest scores.

	Low Error Group			Medium Error Group			High Error Group		
	Polite Mean (sd)	Direct Mean (sd)	Cohen's d	Polite Mean (sd)	Direct Mean (sd)	Cohen's d	Polite Mean (sd)	Direct Mean (sd)	Cohen's d
Immediate Posttest	0.66 (0.21)	0.68 (0.19)	−0.08	0.64 (0.22)	0.59 (0.21)	0.22	0.63 (0.19)	0.50 (0.24)	0.66*
Delayed Posttest	0.67 (0.18)	0.66 (0.18)	0.06	0.65 (0.20)	0.62 (0.18)	0.16	0.63 (0.17)	0.57 (0.20)	0.36

*Significant difference, $p < 0.05$.

analysis revealed that the High Error group performed significantly better on the immediate posttest with the polite tutor than with the direct tutor ($d = 0.66$; $p = 0.04$), but this significant difference did not occur for either the Medium or the Low Error groups ($ps > 0.7$). The effect on the immediate posttest of the High Error group is due to the near transfer problems of the posttest, with the polite tutor group performing significantly better than the direct tutor group for High Error students ($p = 0.03$). In contrast, the Medium and Low Error students did not differ significantly between the polite and direct groups on near transfer problems ($ps > 0.5$). For the conceptual problems, within each Error group, the polite tutor group and the direct tutor group performed similarly ($ps > 0.2$). For the delayed test, no significant main effects of politeness and no significant interactions of politeness with error rate were found, and no significant simple main effects within Error groups were found.

6. Discussion

Did students in classrooms learn more by using the polite web-based tutor than by using the direct web-based tutor? On the one hand, the results of Table 5, discussed above, indicate that the polite tutor did not benefit learning, as compared to the direct tutor, either for low or high prior knowledge learners. As assessed by a variety of dependent variables, there were no significant differences between the polite and direct groups. On the other hand, the results of Table 6 at least hint toward a possible benefit of the polite tutor. In particular, for students who made many errors during the intervention, arguably a low prior knowledge group, the polite tutor ultimately led to more learning as compared to the direct tutor, at least with respect to the immediate posttest near transfer scores (which led, in turn, to the significant overall difference shown in Table 6). In summary, the experiment discussed in this paper provides weak evidence that the polite tutor made a difference to classroom learning of stoichiometry, in particular with respect to those students who had less knowledge to begin with.

How do these results square with prior results on the politeness effect, both our own and those of other researchers? The strongest results in support of the politeness effect come from the McLaren et al. (in press) and Wang et al. (2008) lab studies, both of which provide solid evidence pointing in the direction of polite tutoring making a difference to low prior knowledge students. Using the same tutors, materials, and experimental design reported in the present study, yet applied in a lab setting, McLaren et al. (in press) found that low prior knowledge university students benefitted more from the polite than the direct tutor on both immediate and delayed posttests, and that high prior knowledge students seemed to be adversely affected by the polite tutor. Likewise, Wang et al. (2008) found that non-engineering students in an engineering domain (i.e., low prior knowledge learners) benefitted more than engineering students from using a polite tutor to learn than from using a direct tutor. While the results in the present study are clearly less decisive than either of these prior studies – for instance, none of the dependent variables that demonstrated learning effects in favor of the polite tutor for low prior knowledge subjects in the McLaren et al. (in press) study showed such effects in the present study – they nonetheless provide some indication that low prior knowledge learners stand to benefit the most from polite tutoring. The weaker results of the present study are most similar to the findings of Wang and Johnson (2008). While Wang and Johnson cite evidence for the politeness principle in their lab study involving a web-based tutoring system for teaching foreign language, they actually found support for politeness only on a single dependent measure, utterance formation questions. On a number of other dependent measures, including overall score, multiple-choice questions, and phrase-matching questions, the polite tutor did not have an effect. Likewise, as discussed above, the present study found support for the politeness effect only on a single dependent measure, amongst a number of measures that were evaluated. Finally, the study of McLaren et al. (2007), which was also conducted in a classroom setting with virtually the same materials and design described in the present study, did not reveal a politeness effect. However, McLaren et al. (2007) did not evaluate low versus high prior knowledge learners in their analysis. Furthermore, a post-questionnaire administered as part of this study indicated that both lower confidence and lower prior knowledge students enjoyed working with and were more sensitive to the polite tutor than the direct tutor. Thus, even in this study with no empirical support of the politeness principle, there was at least a hint that more needy students might stand to benefit from polite tutoring.

In summary, it appears that while lab studies have more clearly revealed the benefits of polite tutoring for lower prior knowledge students, the classroom studies have only weakly pointed in this direction. But why is this? Why don't benefits that are clearly observed in lab settings transfer to classroom settings? One possible explanation is that when students are using the tutor as part of an actual science class, as they were in both the present study and the McLaren et al. (2007) study, in which they are motivated to get a good class grade both on the specific materials (since the posttests were used as class or extra credit grades) and in the class in general, those factors determine their level of motivation and performance more than politeness does. As an indication of this, students in one of the Advanced Placement (AP) chemistry classes in the study scored particularly high on the delayed posttest. This suggests that these students, once exposed to the materials and an initial posttest, made a special effort to study for the delayed test. In contrast, the university psychology students in the McLaren et al. (in press) study almost surely did not do this, given no motivation to do so. Another possible factor in the different results across classroom and lab is the most fundamental difference between the two settings: In the rough-and-tumble of the classroom, with its noise, question-asking, and social environment, students may simply not concentrate as much on the feedback provided by the computer tutor. The lab setting, on the other hand, is a quiet environment where subjects work on their own with few distractions, and certainly none from classmates and a teacher. The classroom distractions might lead to the need for more distinctive, attention-grabbing feedback in that setting versus the lab. Finally, the 60–120 min of exposure to the intervention may not be long enough in classroom use to elicit the effect.

Finally, while the present study was not designed to test empirically the use of web-based tutors versus non-web-based tutors, and the application of the politeness principle within each environment, it nevertheless provides some general evidence of the effectiveness of web-based materials. Students in both the polite and direct conditions, all of who used web-based tutors, benefitted from their use. Furthermore, from a practical point of view, the study itself would have been much more difficult to conduct without web-based materials and tutors, as teachers in widely dispersed geographic regions used the materials in their classrooms simply by having their students traverse to a given URL in a web browser. Thus, our third goal in conducting these studies – testing whether we could effectively deliver polite tutors over the web – has largely been demonstrated both in this and earlier studies.

7. Conclusion

Empirically, the study reported in this paper, which was a follow-up to the lab study of (McLaren et al, in press), demonstrated that lower prior knowledge students could benefit more from polite than direct tutoring. However, the study also demonstrated that, like earlier research (McLaren et al., 2007), the politeness effect does not operate as decisively in the classroom as in the lab. While the politeness effect *was* identified for lower prior knowledge students, the effect was not seen according to the same analyses of the earlier, similar lab study and was only detected for a single dependent variable, according to a different definition of lower prior knowledge students than was used in the lab study. Thus, we interpret the politeness effect to be weaker in the present classroom study than it was in the lab study.

While only providing weak support for the politeness effect, this study is nevertheless consistent with social agency theory. That is, at least some students in the study appeared to try harder to understand tutor feedback and to learn when they viewed the tutor as a polite, social partner. In particular, lower prior knowledge students – in the case of the present study, those students who made the most errors during the intervention – seemed to respond better to the polite tutor's feedback and learn more than the students exposed to the direct tutor's feedback. Our hypothesis, consistent with social agency theory, is that politeness tends to engender and encourage generative processing in low prior knowledge learners. On the other hand, high prior knowledge learners may simply have better generative cognitive skills to begin with, allowing them to access and integrate prior knowledge more readily. In the prior lab study (McLaren et al., in press) the polite feedback seemed to actually *hurt* high prior knowledge learners – perhaps as a distraction – but we did not see evidence of that in the present study.

In general, our findings point to the need for intelligent tutor designers to pay attention to how feedback is provided to learners, in addition to what and when it is provided. Intelligent tutors should be developed to have the capability to engage students in a more polite, social manner, as well as to detect when it is and is not appropriate to take such an approach. Given the results we've found both in this and earlier studies, the approach sometimes taken with worked examples, in which examples are initially given to students and then gradually faded (Schwonke et al., 2007), would be worth investigating in the context of politeness. In other words, students may initially benefit from the polite feedback and social engagement from a tutor but, as their generative processing skills in a particular domain improve, may be better off receiving more concise and directive feedback, as suggested by the expertise reversal effect (Kalyuga, 2005). Given the long established ITS subfield of student modeling (Beck & Woolf, 1998; Murray, 1998), including student models of affect (D'Mello, Craig, Witherspoon, McDaniel, & Graesser, 2008; Woolf et al., 2009), it should be within reach of tutoring systems to dynamically adapt their feedback to the knowledge and affective characteristics of the learner. This is one of the key areas we intend to pursue in future work: After establishing the boundary conditions of the politeness principle, as well as determining which students will most likely benefit, we will capture and use this information in real-time to provide polite feedback, as appropriate.

There are two other key directions we will pursue in the future. First, we would like to investigate how much effort students put into trying to make sense of a tutor's feedback. Since the theory of politeness, as embodied in social agency theory, posits that generative processing is enhanced through a student's effort at relating to and understanding a tutor's feedback, it is important to get at just how much effort students expend in learning. For this, we will take a cue from (Paas, 1992) in which effort questions (i.e., prompts asking students to rate their current level of effort) were strategically placed within materials to determine the relative effort expended on different problems between conditions. Second, and most important to the main theme of the present paper, we plan to further investigate the differences in the politeness effect in the classroom versus the lab. One way to investigate this issue would be to simulate more realistic classroom situations in the lab, such as having one condition in which multiple students who know one another, and who are perhaps studying the content domain, work with intelligent tutors at the same time, in the same room. These students could then be compared more directly to participants in a tightly controlled lab environment.

References

Aleven, V., McLaren, B. M., & Sewall, J. (2009). Scaling up programming by demonstration for intelligent tutoring systems development: an open-access website for middle school mathematics learning. *IEEE Transactions on Learning Technologies, 2*(2), 64–78.

Aleven, V., McLaren, B. M., Sewall, J., & Koedinger, K. R. (2009). A new paradigm for intelligent tutoring systems: example-tracing tutors. *International Journal of Artificial Intelligence in Education, 19*(2), 105–154.

Alpert, S. R., Singley, M. K., & Fairweather, P. G. (1999). Deploying Intelligent tutors on the web: an architecture and an example. *International Journal of Artificial Intelligence in Education, 10*(2), 183–197.

Baker, R., Walonoski, J., Heffernan, N., Roll, I., Corbett, A., & Koedinger, K. (2008). Why students engage in "Gaming the System" behavior in interactive learning environments. *Journal of Interactive Learning Research, 19*(2), 185–224.

Beal, C. R., Walles, R., Arroyo, I., & Woolf, B. P. (2007). Online tutoring for math achievement: a controlled evaluation. *Journal of Interactive Online Learning, 6*, 43–55.

Beck, J. E., & Woolf, B. P. (1998). Using a learning agent with a student model. In B. P. Goettl, H. M. Halff, C. L. Redfield, & V. J. Shute (Eds.), *Proceedings of the 4th International Conference on intelligent tutoring systems* (pp. 6–15). Springer.

Brown, P., & Levinson, S. C. (1987). *Politeness: Some universals in language use*. New York: Cambridge University Press.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

D'Mello, S. K., Craig, S. D., Witherspoon, A. W., McDaniel, B. T., & Graesser, A. C. (2008). Automatic detection of learner affect from conversational cues. *User Modeling and User-Adapted Interaction, 18*(1–2), 45–80.

Grice, H. P. (1975). Logic and conversation. In P. Cole, & J. Morgan (Eds.), *Syntax and semantics, Vol. 3* (pp. 41–53). New York: Academic Press.

Hattie, J. A. C., & Gan, M. (2011). Instruction based on feedback. In R. Mayer, & P. Alexander (Eds.), *Handbook of research on learning and instruction* (pp. 249–271). New York: Routledge.

Johnson, C. I., & Mayer, R. E. (2009). A testing effect with multimedia learning. *Journal of Educational Psychology, 101*, 621–629.

Kalyuga, S. (2005). Prior knowledge principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 325–338). New York: Cambridge University Press.
Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: a historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin, 119*, 254–284.
Koedinger, K. R., & Corbett, A. (2006). Cognitive tutors: technology bringing learning science to the classroom. In R. K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (pp. 61–77). New York: Cambridge University Press.
Mayer, R. E. (2005a). Principles of multimedia learning based on social cues: personalization, voice, and image principles. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 201–212). New York: Cambridge University Press.
Mayer, R. E. (2005b). Principles for reducing extraneous processing in multimedia learning: coherence, signaling, redundancy, spatial contiguity, and temporal contiguity. In R. E. Mayer (Ed.), *Cambridge handbook of multimedia learning* (pp. 183–200). New York: Cambridge University Press.
Mayer, R. E. (2009). *Multimedia learning* (2nd ed.). New York: Cambridge University Press.
Mayer, R. E., Johnson, L., Shaw, E., & Sandhu, S. (2006). Constructing computer-based tutors that are socially sensitive: politeness in educational software. *International Journal of Human Computer Studies, 64*, 36–42.
McLaren, B. M., Lim, S., Yaron, D., & Koedinger, K. R. (2007). Can a polite intelligent tutoring system lead to improved learning outside of the lab? In R. Luckin, K. R. Koedinger, & J. Greer (Eds.), *Proceedings of the 13th International Conference on Artificial Intelligence in Education (AIED-07), Artificial intelligence in education: Building technology rich learning contexts that work* (pp. 433–440) Amsterdam: IOS Press.
McLaren, B. M., Lim, S., & Koedinger, K. R. (2008). When and how often should worked examples be given to students? New results and a summary of the current state of research. In B. C. Love, K. McRae, & V. M. Sloutsky (Eds.), *Proceedings of the 30th annual conference of the cognitive science society* (pp. 2176–2181). Austin, TX: Cognitive Science Society.
McLaren, B. M., DeLeeuw, K. E., & Mayer, R. E. (in press). A politeness effect in learning with web-based intelligent tutors. To appear in the *International Journal of Human Computer Studies*.
Melis, E., Andrès, E., Büdenbender, J., Frischauf, A., Goguadze, G., Libbrecht, P., et al. (2001). ActiveMath: a generic and adaptive web-based learning environment. *International Journal of Artificial Intelligence in Education, 12*, 385–407.
Mostow, J., & Beck, J. (2007). When the rubber meets the road: lessons from the in-school adventures of an automated Reading Tutor that listens. In B. Schneider, & S.-K. McDonald (Eds.), *Conceptualizing Scale-Up: Multidisciplinary Perspectives, Vol. 2* (pp. 183–200). Lanham, MD: Rowman & Littlefield.
Murray, W. R. (1998). A practical approach to Bayesian student modeling. In B. P. Goettl, H. M. Halff, C. L. Redfield, & V. J. Shute (Eds.), *Proceedings of the 4th International Conference on intelligence tutoring systems* (pp. 424–433). Springer.
Nass, C., & Brave, S. (2005). *Wired for speech*. Cambridge, MA: MIT Press.
Paas, F. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: a cognitive load approach. *Journal of Educational Psychology, 84*, 429–434.
Person, N. K., Kreuz, R. J., Zwaan, R. A., & Graesser, A. C. (1995). Pragmatics and pedagogy: conversational rules and politeness strategies may inhibit effective tutoring. *Cognition and Instruction, 13*, 161–188.
Reeves, B., & Nass, C. (1996). *The media equation*. New York: Cambridge University Press.
Rickel, J., & Johnson, W. L. (1999). Animated agents for procedural training in virtual reality: perception, cognition, and motor control. *Applied Artificial Intelligence, 13*, 343–382.
Roediger, H. L., III, & Karpicke, J. D. (2006). The power of testing memory: basic research and implications for educational practice. *Perspectives on Psychological Science, 1*(3), 181–255.
Schwonke, R., Wittwer, J., Aleven, V., Salden, R., Krieg, C., & Renkl, A. (2007). Can tutored problem solving benefit from faded worked-out examples? In S. Vosniadou, D. Kayser, & A. Protopapas (Eds.), *Proceedings of EuroCogSci 07. The European cognitive science Conference 2007* (pp. 59–64) New York, NY: Erlbaum.
Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research, 78*(1), 153–189.
VanLehn, K. (2006). The behavior of tutoring systems. *International Journal of Artificial Intelligence in Education, 16*(3), 227–265.
VanLehn, K., Lynch, C., Schulze, K., Shapiro, J. A., Shelby, R., Taylor, L., et al. (2005). The Andes physics tutoring system: lessons learned. *International Journal of Artificial Intelligence and Education, 15*(3), 1–47.
Wang, N., & Johnson, W. L. (2008). The politeness effect in an intelligent foreign language tutoring system. In B. P. Woolf, E. Aimeur, R. Nkambou, & S. Lajoie (Eds.), *Proceedings of the 9th international conference on intelligent tutoring systems, Lecture notes in computer science, Vol. 5091* (pp. 260–280). Berlin: Springer.
Wang, N., Johnson, W. L., Mayer, R. E., Rizzo, P., Shaw, E., & Collins, H. (2008). The politeness effect: pedagogical agents and learning outcomes. *International Journal of Human Computer Studies, 66*, 96–112.
Wood, H., & Wood, D. (1999). Help seeking, learning, and contingent tutoring. *Computers and Education, 33*, 153–169.
Woolf, B. P., Burleson, W., Arroyo, I., Dragon, T., Cooper, D. G., & Picard, R. W. (2009). Affect-aware tutors: recognizing and responding to student affect. *International Journal of Learning Technology, 4*(3/4), 129–164.