2019

The effectiveness of using intelligent tutoring systems to increase student achievement

Tedi Swanson
University of Northern Iowa

The effectiveness of using intelligent tutoring systems to increase student achievement

Abstract

Intelligent Tutoring Systems could be used to provide differentiated instruction. This review examines qualities of Intelligent Tutoring Systems and their impact on student achievement. Thirty peer-reviewed research studies published from 1997 to 2019 were selected for analysis. This review considers how intelligent tutoring systems compare with other methods of instruction, and how an intelligent tutoring system's on-screen tutor impacts student achievement. Finally, this review considers methods of ITS personalization and how those methods impact student achievement. The reviewed research studies indicated that ITS was more effective than all forms of instruction except small group and individualized instruction. Additionally, on-screen agents in and personalization of Intelligent Tutoring Systems often have a positive impact on student learning. Recommendations for classroom implementation of intelligent tutoring systems and suggestions for future research are discussed

The Effectiveness of Using Intelligent Tutoring Systems to Increase Student Achievement

A Graduate Review

Submitted to the

Division of Instructional Technology

Department of Curriculum and Instruction

In Partial Fulfillment

Of the Requirements for the Degree

Master of Arts

UNIVERSITY OF NORTHERN IOWA

by

Tedi Swanson

May, 2019

EFFECTIVENESS OF INTELLIGENT TUTORING SYSTEMS

This Review by: Tedi Swanson

Titled: The Effectiveness of Using Intelligent Tutoring Systems to Increase Student Achievement

has been approved as meeting the research requirement for the Degree of Master of Arts.

___7/27/2019___ _________Leigh Zeitz______________________
Date Approved Graduate Faculty Reader

___7/27/2019__ _________Deborah Tidwell__________________
Date Approved Graduate Faculty Reader

___7/27/2019__ ________Dr. Robin Dada___________________
Date Approved Head, Department of Curriculum and Instruction

Abstract

Intelligent Tutoring Systems could be used to provide differentiated instruction. This review examines qualities of Intelligent Tutoring Systems and their impact on student achievement. Thirty peer-reviewed research studies published from 1997 to 2019 were selected for analysis. This review considers how intelligent tutoring systems compare with other methods of instruction, and how an intelligent tutoring system's on-screen tutor impacts student achievement. Finally, this review considers methods of ITS personalization and how those methods impact student achievement. The reviewed research studies indicated that ITS was more effective than all forms of instruction except small group and individualized instruction. Additionally, on-screen agents in and personalization of Intelligent Tutoring Systems often have a positive impact on student learning. Recommendations for classroom implementation of intelligent tutoring systems and suggestions for future research are discussed.

keywords: Intelligent Tutoring system, student achievement

Table of Contents

The Effectiveness of Intelligent Tutoring Systems to Differentiate Instruction

Introduction

Twenty-seven students. One English Language Learner. Four students who perform significantly below grade level, at the 4th percentile or below. Three students who perform significantly above grade level, at the 88th percentile or above. One student who performs at the 97th percentile. The remaining eighteen students hover in varying degrees of closeness to proficiency. This is a snapshot of a typical public school classroom (Rock, Gregg, Ellis, & Gable, 2008).

Differentiation, as defined by Merriam-Webster dictionary (2019), is the process of developing distinguishing characteristics. In this case, it is the process of developing distinguishing forms of classroom instruction to better match learning to the zone of proximal development (ZPD) of any given student. It is nearly impossible for a single teacher to meet the needs of wildly varied groups of students, such as the one described above, and yet such differentiation is essential for quality education. The use of effective Intelligent Tutoring Systems could bridge this gap.

An Intelligent Tutoring System (ITS) is a computer tutoring system that utilizes artificial intelligence and cognitive theory to guide learners through problems and solutions “by creating hints and feedback as needed from expert-knowledge databases” (Kulik & Fletcher, 2015, p. 43). The purpose of this literature review is to analyze the research available on Intelligent Tutoring Systems and their effectiveness within classrooms serving elementary to graduate-level students. To be deemed effective, the use of the ITS needs to result in significant gains in student achievement.

Themes that emerged while reviewing the literature were:

1. The effectiveness of Intelligent Tutoring Systems in comparison with other methods of instruction.
2. The impact of the ITS on-screen tutor (animated pedagogical agent) on student achievement.
3. Effect of various methods of ITS personalization on student achievement.

This review will analyze intelligent tutoring systems in secondary education classrooms. In order to use Intelligent Tutoring Systems in the classroom as a method of differentiation, it is essential to identify when and how they are the most effective.

The results of this review are a starting point for identifying the qualities and applications of Intelligent Tutoring Systems that have the most impact on student learning. The assessment could provide context for educators searching for ways to bolster the differentiation in their classrooms and will suggest methods for application and implementation of Intelligent Tutoring Systems that will provide the most meaningful experience for students and will have an impact on student achievement.

Methodology

Locating valid and reliable sources on the topic of Intelligent Tutoring Systems was a challenge. I believe this is because the field of Intelligent Tutoring Systems is still in its early stages. Intelligent Tutoring Systems are still in the process of becoming widespread and common practice. To locate the available sources, I began by using the term *Intelligent Tutoring System* in OneSearch, Ebsco ERIC, and Google Scholar. These databases were all accessed through the University of Northern Iowa library which provided the ability to filter articles based on criteria for reliability and validity. The first literature review I found that was relevant to the topic was entitled "Effectiveness of Intelligent Tutoring Systems" so I added *effective* to the search terms. Upon further consideration, both *differentiation* and *achievement* were also added. After finding several articles, I then identified the source of those articles. I reviewed journals such as *Computers & Education* and the *International Journal of Human-Computer Studies* for articles relevant to my topic. I did this by looking at the table of contents and also by searching for "intelligent tutoring system" in the journal's database. Once I located relevant articles, I typed the titles of those articles into OneSearch in order to locate full text of them. I also used sources referenced in the articles I found relevant to my research as additional sources.

Upon receiving results from these searches, the articles I chose to analyze evaluated or performed research on an aspect (i.e. data taken into account when formulating the ITS' feedback) of an intelligent tutoring system, discussed how that aspect of the ITS impacted student achievement, and was from a peer-reviewed source. My analysis of these articles involved identifying the author and the authors' affiliations, and ensuring the article included quantitative data about the effectiveness of the ITS on student achievement. Learner age level was not a restriction that I took into account as research at all educational levels was beneficial to understanding the effectiveness of ITS.

Articles I ultimately chose to include in this review provided significant findings about aspects of Intelligent Tutoring Systems and their level of impact on student achievement. The articles were all published within the last seven years, aside from one article that provided foundational information about the origins of Intelligent Tutoring Systems. The primary goal of this review was to explore the effects of Intelligent

Tutoring Systems and the teaching strategies that they use on student achievement. All articles included in this review were selected to provide a richer understanding of those qualities that had the biggest impact on student achievement.

Analysis and Discussion

Intelligent Tutoring Systems (ITS) are known for providing individualized and differentiated instruction based on student need (Ma, Adesope, Nesbit, & Liu, 2014). These programs are responsive to student input and provide immediate feedback to students to further deepen their understanding and improve their learning. Programs like IXL (IXL Learning, 2019) and Prodigy (SMARTeacher INC., 2011) first require students to complete a diagnostic test to identify students' ability levels and learning gaps, and then provide instruction to students based on the results of those diagnostics. With an ever-growing need for differentiation in the classroom, these programs are able to adapt instruction to students in a way that an individual classroom teacher cannot. How effective are these ITS programs in increasing student performances? Do certain qualities of an ITS make it more or less effective in increasing student achievement?

This review will consider how ITS compares with other methods of instruction in terms of resulting student performance; the impact of the ITS on-screen tutor (animated pedagogical agents) on student achievement; and the effects of various methods of ITS personalization on student achievement.

ITS in comparison with other methods of instruction.

When considering the implementation of Intelligent Tutoring Systems in classrooms, the first question that demands to be asked is how Intelligent Tutoring Systems compare to other methods of instruction. Can a computer provide better instruction than a human being? Can a computer truly have a positive impact on student achievement? Kulik and Fletcher (2016), and Steenbergen-Hu and Cooper (2014) found that intelligent tutoring systems had a more positive impact on student learning than conventional classes. Ma, et al. (2014) similarly found that ITS had a more positive impact on student achievement than other methods of instruction, except for small group instruction and one-on-one human tutoring. Holdich and Chung (2003), Mostow, Nelson-Taylor, and Beck (2013), Xin, Tzur, Hord, Liu, Park, and Si (2017) conducted students analyzing the use of ITS instruction in writing, reading, and mathematics respectively in comparison with traditional classroom instruction and found ITS instruction to have positive effects on student achievement in these specific situations. Martin, Klein, and Sullivan (2006) and Sosa, Berger, Saw, and Mary (2011) posited that ITS made a significant impact on student learning because it provided more opportunities for practice and feedback, while Martin et al. posited that ITS made a significant impact on student learning because it provided more opportunities for practice and feedback. Sosa et al., additionally found that prior knowledge has some impact on student achievement as a result of ITS instruction. Ultimately, findings show that while individualized tutoring is the ideal form of instruction, Intelligent Tutoring Systems provide instruction that has a significant impact on student learning over traditional classroom instruction.

Significance of ITS instruction on student performance. In a meta-analysis, Kulik and Fletcher (2016) compiled 50 studies on ITS with the stipulation that all studies they analyzed involved a control group that received ITS instruction. They distinguished ITS instruction from other computer-based instruction by requiring that the ITS operated from a knowledge database and that they used a computational and dialogue-generating tool that extracted relevant information from said knowledge database. The final collection of studies used for analysis covered instruction in elementary schools, high schools, colleges, and military training institutions. The ITS instruction provided in these studies ranged from one hour to 48 weeks of instruction. After analyzing these 50 studies, Kulik and Fletcher (2016) discovered that in 46 of the 50 studies, students who received ITS instruction outperformed students in traditional classrooms. The median student achievement growth in the 50 students was equivalent to a test performance growth from the 50th to 75th percentile. Kulik and Fletcher also reported that increased student achievement was more significant on local tests than on standardized tests, although test growth was still evident on standardized tests.

Steenbergen-Hu and Cooper (2014) conducted a meta-analysis of ITS impact on college students' achievement. For the purpose of the analysis, they defined an ITS as "highly adaptive, interactive, and learner-paced learning environments operated through computers" (p. 331) They also required that a program be a domain-related stand-alone computer tutorial in order for the study to be included in the meta-analysis. The 26 studies used in the analysis compared ITS instruction with conventional instruction, computerized instruction that was not deemed an ITS, and self-reliant learning, such as homework. Steenbergen-Hu and Cooper found that ITS instruction had a more significant impact on student achievement in comparison with all other instruction except human tutoring, with an average effect size of .32. Steenbergen-Hu and Cooper's analysis further confirms that ITS instruction can have a positive impact on student achievement in relation to other methods of instruction.

Ma et al., (2014) conducted a similar meta-analysis in which they defined an ITS as a system that performs tutoring functions and draws conclusions from student responses in order to track student understanding. Ma et al. found that most studies compared the use of ITS with large-group human instruction, individual computer-based instruction that differed from ITS based on the definition, and the individual use of textbooks or workbooks. The meta-analysis showed that ITS produced moderate effect sizes in comparison with these methods of instruction, ranging from .36 to .57. Additionally, they found that small group instruction of 8 students or fewer and individual human instruction provided a small non-statistically significant advantage over ITS instruction ($p < .001$).

Holdich and Chung (2003) conducted a study using an ITS tutor that provided narrative writing instruction to students. Understanding that providing feedback and guidance for student writing can be difficult for teachers due to time limitations and large class sizes, this ITS, named HARRY, was developed to help meet the writing needs of students. Within a class of third graders, a control group that did not receive instruction from HARRY was compared with a group of students who wrote one story without the assistance of HARRY and one within the ITS assistance. HARRY provided guidance for students throughout the composition process by reminding them of the qualities of good writing and what kind of details or information they should include. HARRY also provided editing and revision feedback. Students' writing was then compared based on organization and grammar. Findings showed that students who interacted with HARRY and implemented feedback from the system improved on their second story, while students who did not use HARRY did not improve. However, Holdich and Chung acknowledge that this improvement requires students to fully interact with the feedback as they are using the ITS system, and that the HARRY system provides heavy scaffolding for students with the goal that said scaffolding would be reduced over time and eventually removed.

Similarly, the ability of a teacher to conduct guided reading with individual students in a standard classroom becomes difficult with large class sizes. Mostow et al. (2013) conducted a study that compared ITS-based guided oral reading against sustained silent reading (SSR) in grades 1-4. One hundred and seventy eight students participated in the study. Using a reading assessment, students were put in pairs based on reading ability level. Then one student within a pair was randomly assigned to use the ITS, while the other student was assigned to the SSR group. The ITS for guided oral reading used speech detection to provide prompting to students and identify mispronounced words, although Mostow et al. identify that the tutor was insufficient at identifying miscues or situations when students would say a correct word but the word was not present in the reading passage. Students were then given a post test on reading ability. Students in the SSR group never outperformed students in the ITS group. Students in the ITS group showed significant gains in the area of word identification in comparison to students in the SSR group. Additionally, the ITS group showed gains in word blending, and spelling. Although the use of the ITS did not cause widespread gains in student achievement, the areas the tutor could effectively target (e.g., word identification skills, such as phonemic awareness and word blending) resulted in significant gains in student achievement.

Xin, Tzur, Hord, Liu, Park, and Si (2017) analyzed the use of a mathematics ITS with students with learning difficulties. The study was conducted with 17 elementary students who were randomly assigned to either a teacher-led group, or a group that used the ITS. Results were determined using a pre and post test. The results found that although both the teacher-led instruction and ITS instruction resulted in improved student achievement, the students using the ITS saw increased student achievement more quickly than those students receiving the teacher-led instruction.

Ultimately, these meta-analyses and studies have confirmed that small group and individual instruction continue to outperform ITS. In comparison with most other forms of instruction, however, ITS had a statistically significant positive impact on student achievement. This was especially true in areas of instruction, such as reading and writing, that are increasingly difficult for teachers to address effectively as class sizes increase. Additionally findings from the study conducted by Xin, et al. show that when in direct comparison, ITS based instruction could result in increased student achievement at a quicker rate than whole class teacher-led instruction.

ITS instructional elements having a significant impact. Sosa et al. (2011) performed an analysis of 45 studies that reported "empirical outcome evaluation of computer-assisted instruction in statistics in comparison to a control condition." (p. 101) The studies analyzed the use of ITS in statistics instruction. The results from the meta-analysis showed that in comparison with lecture-only instruction, ITS had a significant impact on student achievement. With an effect size of 0.33, 63% of students receiving ITS instruction outperformed the control group who did not receive ITS instruction. They found that this was especially true when the group receiving ITS instruction received more overall instruction than the control group receiving lecture-only instruction. Although this seems like an obvious conclusion, the results capitalize on the idea that ITS instruction allows for more practice with material and therefore higher student achievement than lecture-only instruction. Sosa et al., also found that the computer-assisted instruction had a bigger impact on graduate students than on undergraduate students, indicating that prior-knowledge about the content being presented had an impact on how effective the computer based-instruction was in improving student achievement.

Martin et al. (2006) conducted a study with 256 undergraduate students in a computer literacy course. They were given six-different versions of a lesson that included various combinations of instructional elements: information, objectives, practice with feedback, examples, and review. The variations ranged from a program with information only to a full program with all five instructional elements. Results from the study were measured - using a pre and post test and an attitudinal survey. The mean scores from the pre to post test indicated that practice had the most significant impact on student achievement. Students receiving the full program achieved a mean score of 17.61 out of 20 while students whose program did not have practice received a mean score of 14.98 out of 20. All other programs with elements removed impacted the mean score by less than one point, with the removal of examples having the second-biggest impact on student achievement and lowering the median score by .45. Martin et al. concluded, similarly to Sosa et al. (2011), that the additional practice afforded in an ITS has the biggest impact on student achievement.

Impact of the On-Screen Tutor

One aspect of Intelligent Tutoring Systems that has been examined is the role that the on-screen tutor, or animated pedagogical agent, plays in improving student achievement. Mayer, Johnson, Shaw, and Shandu (2006) indicated that these agents could be significant in providing feedback to the learner through suggestions for next steps, corrections, and explanations. Moreno and Flowerday (2006) discovered that when given a choice, learners preferred to work with on-screen agents. Across all research, it was widely hypothesized that the on-screen agent would have some positive impact on learner achievement. Frechette and Moreno (2010) and Cook, Friedman, Duggan, Cui, and Popescu (2016) analyzed the agent's use of hand gestures and facial expressions and the resulting student achievement. Moreno and Flowerday (2006) and Kim (2016) considered the effect that the physical appearance of the agent had on student learning experiences. Finally, Mayer, et al. (2005) and McLaren, DeLeeuw, and Mayer (2010) examined how agent politeness impacted student achievement. Synthesizing the findings of these researchers provides a rich understanding of the impact of animated pedagogical agents on student achievement.

Agent hand gestures and facial expressions. Frechette and Moreno (2010) conducted research to study the impact of facial expressions and gestures of an on-screen agent. The study was conducted with 93 college students. Students were randomly assigned to a control group which included no agent, a static group with a non-animated agent, a deictic group that included a tutor with hand-and-arm gestures, or a fully-animated group wherein the agent had lip-synching, full facial expressions and hand and arm gestures. The deictic group received this nomenclature as a means of signifying that the hand and arm gestures "reference, reflect, or reinforce the verbal information" (p. 62).

The program given to students was a linear multimedia presentation and was followed up with comprehension tests, as well as attitudinal surveys. Results from this study showed that there were no statistically significant differences in comprehension tests between the various groups, indicating that the agent did not have a significant impact on student learning. Results also indicated that the agent with full facial expressions actually lowered student comprehension. Frechette and Moreno (2010), however, did identify a limitation of their study: the agent was nonessential to the learning and did not engage with the learner due to the linear nature of the presentation. They concluded that in similarly linear learning environments, an agent is not beneficial, but the agent may be beneficial in a more interactive and responsive learning environment.

In contrast, Cook, et al. (2016) found hand gestures to be essential in linear lessons on mathematics. Sixty-five children with a median age of nine years old participated in the study. The study used a computer-generated animated avatar standing in front of a virtual whiteboard and explaining a mathematical concept. Pairs of videos were established wherein one avatar in the pair did not use hand gestures, while the other avatar in the pair used hand gestures both to reinforce content and to increase the charisma and appeal of the avatar. The study found that students who received the lesson from the avatar with hand gestures not only learned more from the lessons than those with the gesture-less avatar, but also children who received the instruction from the avatar with gestures were better able to transfer their learning to problems on the posttest than children who learned from the gesture-less avatar.

Agent physical appearance. Moreno and Flowerday (2006) and Kim (2016) considered the physical appearance of the agent. Moreno and Flowerday's study included 80 undergraduate participants. The study involved a choice group that was able to select a pedagogical agent and a non-choice group that was assigned a pedagogical agent. The study hypothesized that learners would both choose more often and learn more from pedagogical agents that were perceived as representing them. The results of the study indicated that the choice of a same-gender or same-ethnicity agent had no significant impact on student learning. Although the impact was not statistically significant, two results were of note: students of color were more likely to choose an agent that represented them; Additionally, students who chose to learn from a different-ethnicity agent outperformed those who chose an agent of the same ethnicity. They also outperformed students who were assigned a different-ethnicity agent. The researchers believe these findings show both support for the benefit of student choice and support for learner distraction as the students who chose same-ethnicity agents may have been focusing more on whether the agent represented them than on the materials being presented.

Kim's (2016) study had similar results. The study was focused specifically on whether agent representation could encourage female learners in a mathematics course both academically and emotionally. The study included 67 female students in a ninth-grade introductory algebra course. Learners were randomly assigned to one of four pedagogical agents: female peer, male peer, female teacher or male teacher. Learners were given a questionnaire about their perception of the agent, pre and post questionnaires to determine their attitudes toward mathematics, and pre and post tests over the mathematical concepts covered by the tutor. The results of the study showed that both ethnic-minority girls and Caucasian girls felt more positively about peer-aged agents regardless of gender. For ethnic-minority girls, they viewed the male peer agent more positively than the female peer agent. The researchers believe this difference may be due to the female peer creating a feeling of competitiveness in the learners. Additionally, while both groups felt more positively about the peer agent, Caucasian girls learned more from the teacher agent and ethnic-minority girls learned more from the peer agent. The researcher concluded that the agent had a distinct social role in both the girls' attitude toward learning and the learning outcomes themselves.

Agent politeness. Mayer, et al. (2005) conducted a study to determine whether learners were sensitive to the politeness of feedback given from an on-screen agent. The study examined five politeness strategies and how these strategies impacted student achievement. The participants in the study were 47 college students. The students were given two politeness rating documents and instructed to pretend that the examples they were given were taking place in a computer simulation game. The findings showed that learners were sensitive to differences in politeness. The findings also showed that students who used computers less frequently were more sensitive to the politeness of the tutor's statements than those who used computers more often. The researchers believe the results support the importance of developing socially sensitive agents.

Similarly, McLaren et al. (2010) conducted a study to examine when agent politeness was most essential to learners. There were 132 high school participants in the study that were randomly assigned to the type of feedback they received, either polite feedback or direct feedback. Before the study, students were given an assessment to determine their level of prior knowledge. Students then participated in a Stoichiometry lesson through a web-based program that provided them with both polite and direct feedback. This study did not result in any significant findings in favor of either the polite or direct tutor. There was, however, some weak evidence that students with lower prior knowledge benefitted from the polite tutor and that students put more effort into understanding the feedback from the polite tutor. The researchers believe their results may indicate that as students are processing new learning, a polite tutor may be beneficial, but as the learner becomes more advanced, direct feedback is the most efficient and effective.

From these various studies, it is clear that pedagogical agents are not beneficial in linear learning environments, and that in these linear learning environments, fully-animated pedagogical agents can be distracting (Frechette & Moreno, 2010). However, hand and arm gestures do have an impact on student learning when used to complement the material being presented (Cook et al., 2016). In non-linear learning environments, the gender, age, and ethnicity of the pedagogical agent has no significant impact on the learning of Caucasian students but may have a more significant impact on the learning of minority students (Kim, 2016; Moreno & Flowerday, 2006. Additionally, learners are sensitive to the politeness of agent feedback (Mayer et al., 2005). Polite tutor feedback is viewed more favorably, but is most beneficial to those receiving new learning from the program. For learners that are more advanced in their learning, direct feedback becomes more effective (McLaren et al., 2010). The on-screen agent should be adapted to both the learner's needs and the learning environment in order to have a positive impact on student achievement.

Methods of ITS personalization

One of the defining features of ITS in comparison with computer-based instruction is that it is responsive to the learner and provides adequate feedback and scaffolded instruction based on student understanding. What makes this personalization and responsiveness more or less effective in impacting student achievement? Icoz, Ozdemir, Sanalan, Kaya, and Cakar (2015) examined the use of ontologies to build an ITS and discovered that while effective, the difference between expert and student concept mapping may impact the effectiveness of this method on student achievement. Baker, Corbettt, Koedinger, and Wagner (2004) discovered that without certain system personalizations taken into account with ITS adapting to student performance, students would use the system against itself in order to find the correct answers by guessing until the answer was revealed. Walkington and Bernacki (2019) and Fanscali and Ritter (2014) discovered that personalizing practice within an ITS to student interests increases student engagement but does not always increase student achievement as student engagement with their interests varies and therefore the level of personalization that should be used within the ITS is not universal. Jackson and McNamara (2013) discovered that adapting an ITS into a game-based system similarly peaked student interest and engagement, but did not have a significant impact on student achievement. Finally, Crossley, Allen, Snow, and McNamara (2016) found that combining traditional methods of providing student feedback, such as the use of text feature analysis in automated essay scoring, with individual student information allows for better prediction of student success as well as more effective feedback. The personalization of an ITS allows it to be more effective in positively impacting student achievement.

Personalization of ITS based on student understanding. Kim and Moon (2013) identified the fact that social media tools are impacting the way that students learn and process information, in that students are consuming, producing, and sharing information in networks rather than in the linear format traditionally found in e-learning. Kazi, Haddawy, and Suebnukarn (2010) identified the need for reorganization of ITS to better personalize the ITS feedback to student learning. They identified that traditional ITS are built with a selection of correct answers and if students answers are partially correct do not receive acknowledgement from traditionally organized systems. Kazi et al. recommended the organization of ITS based on ontology in order to counteract this problem. To further these ideas, Icoz et al. (2015) examined the use of ontologies to respond to student understanding within an ITS. An ontology is a way of organizing information about a subject and representing the concepts within that subject and the connections between them. For all intents and purposes, ontologies are commonly known as concept maps. Icoz et al., conducted a study on the use of ontologies to make instructional decisions "based on semantically connected pathways very similar to learner's cognitive structure" (p. 1040). To conduct the study, Icoz et al. had experts compose ontology for a unit on "Our Body and Systems" for a seventh grade science course. They used this ontology to build instruction within an ITS that would respond to student understanding as they progressed through the ITS. The study was conducted with 127 students in five seventh grade science sections. The students were given a paper-based pre and post test. Based on their answers to the first test, students were recommended areas to review and study again before taking the post test. Whether the students completed the recommended studying was not tracked, but answers on the post test increased anywhere from 60 to 100% in the recommended areas. Findings from this study suggest that personalizing learning based on ontology has a positive impact on student achievement, as there were significant gains on the post test when students were given individualized recommendations for learning reinforcement. However, Icoz et al.

also took time to compare students' concept maps based on pre-test results with the ontology crafted by experts and discovered that students' concepts maps differ from that of experts and could result in some gaps in understanding as a result.

Baker et al. (2004) identified necessary ITS personalization in order to ensure that students were using the ITS in way that could impact their achievement. They conducted a study with five middle school classrooms containing seventy students total. Students were given a pre and post test to assess the impact of the ITS on student learning. During the student, Baker et al. analyzed student on and off task behavior. They discovered that one form of off task behavior, referred to as "gaming the system" had a negative impact on student achievement. "Gaming the system" refers to student behaviors in which they randomly select or input answers until the system deems their answer correct without taking into consideration any of the ITS feedback. This behavior results in a -0.34 correlation to the pretest score. As a result, Baker et al., determine it necessary for the design of ITS to be responsive to student understanding in that the design is able to counteract students gaming in the system through methods like mastery learning and eliminating tools which give answers directly to students.

Personalization of ITS based on student interests. Several studies, including Renninger and Su (2012) and Renninger and Hidi (2016) indicate that leveraging student interest has a positive impact on student achievement. Ku and Sullivan (2000) identified a correlation between lower-ability students and increased student achievement when given personalized mathematics problems. Walkington and Bernacki (2019) took personalization to another level in terms of an algebra ITS. They wanted to examine the effect of personalizing story problems to high school student interests and how that impacted student engagement and performance, as students tend to lose interest in subjects like mathematics in adolescence but have high-engagement in out-of-school interests. To conduct this study, they used an ITS with 106 high school students. Students were randomly assigned to a control group, which provided students with standard story problems, a surface personalization group which simply swapped out words in the story problem to match student interests, such as substituting "you are an inspector for a chain of GameStop stores" for "you are a lightbulb inspector" (p. 68). for a student that was interested in video games, or a deep personalization group which completely changed the story problems to match the way that numbers would be naturally used based on the students interest. Walkington and Bernacki identified that one issue arose with randomly assigning students to groups which is that some students have surface-level interest in a topic and were assigned to the deep personalization group, which may have had an adverse effect on the results as these students did not have the background knowledge required to fully comprehend the problem. Ultimately, Walkington and Bernacki found that personalized questions did not cause students to significantly outperform students with standard story problems, but they did result in higher student engagement, as evidenced by a 0.86 increase in correct questions per minute. Overall, there was no difference between students with deep personalization and students with surface personalization, however when looking specifically at students who engage with their interests, deep personalization outperformed surface personalization by 16.5%. Additionally, of the students who have low engagement with their interests, those that received surface personalization outperformed deep personalization by 16.2%.

Similarly, Fanscali and Ritter (2014) examined the impact of personalizing the context of questions within a mathematics ITS. They began by having students rate their interests based on a five-star scale with five stars indicating the most favorable rating. Fanscali and Ritter found that 15.7% of the learners surveyed rated all interests with five stars, and that 25.7% of learners surveyed rated all interests the same. They recognized that this could indicate several possibilities about the learners and took into account those students with "strong" preferences, referring to those students that rate on interest area with five stars and at least one interest area with one or no stars. Fanscali and Ritter also asked students to provide the name of classmates to be inserted into word problems. They found that both students who provided names and students that had a strong preference worked through problems more efficiently than students who did not set preferences, or did not provide names. They were not significant outcomes showing that either of these variables had a significant impact on student achievement. Much like Walkington and Bernacki (2019), Fanscali and Ritter found that this personalization aided in student engagement but did not greatly impact student achievement during the use of the ITS.

In a different vein, Jackson and McNamara (2013) compared the use of a standard ITS and a game-based ITS and their impact on student achievement. The study hypothesized that the game-based system would have a greater positive impact on student achievement due to student interest in the format. The student was conducted with 84 high school students. Student achievement was assessed with a pre and post test, and survey questions throughout to assess students attitude about and motivation to use the two ITS systems. The study was conducted over eleven sessions. During each session, students were randomly assigned to either the game-based system or the standard system.

The results indicated, just as the previous studies found (Fanscali & Ritter, 2014; Walkington & Bernecki, 2019) that both systems showed equivalent increase in student performance, but the game-based system resulted in a more positive student experience.

Personalization of ITS using student information. Johnson, McCarthy, Kopp, Perret, & McNamara (2015) identified one way that ITS could personalize feedback based on student information. W-Pal and iStart ITS provide instruction in writing and reading respectively. These systems provide adaptive feedback based on student performance, but at the time of the report did not provide feedback adaptive to students' developing skills as they use the program. Johnson et al. developed a feature within iStart that allowed questions to adapt to student learning by providing more difficult questions in response to student success. W-Pal has been adapted to provide instruction based on identified weaknesses within a students' writing. Once a student has made revisions based on this targeted area, feedback is adjusted to provide students with a new area of focus. Johnson et al. hypothesize that the addition of these elements will lead to greater increase in student achievement as a result of the adjustments.

Crossley et al. (2016) conducted a study analyzing the impact of learner characteristics taken into account in automatic essay scoring ITS. The study was done using an ITS called The Writing Pal that provides writing feedback to students in the pre-writing, drafting, and revising. Traditionally, automated essay scoring has been done through the use of text features. Crossley et al. examined the use of student information such as demographics, standardized test scores, and survey results, in conjunction with text feature analysis to predict the quality of a persuasive essay within Writing Pal. The study included 86 public high school students, thirty-eight of which identified as English

Language Learners. A pretest essay and a posttest essay were used to determine student achievement. Students used the Writing Pal tutor during the eight sessions in-between the pre and post test essays. Students were randomly assigned to receive feedback based only on student information, only on text features, or on a combination of both. Results found that the combination of student information and text feature analysis was most effective in predicting student success and could result in ITS being able to provide accurate and effective feedback to students. Student information alone was the least effective, as writing is most commonly judge based on text features. All three students indicated that the more personalized an ITS is able to be, the more effective it becomes in improving student achievement.

Conclusions and Recommendations

It is an ongoing trend in the world of education to stress the importance of individualized learning and student-centered classrooms. While no teacher would negate the necessity of meeting the needs and interests of every student in a classroom, the ability of one individual to do so is becoming less and less possible as classrooms become more diverse and class sizes continue to include twenty students or more. Intelligent Tutoring Systems could help solve this problem. Before ITS becomes a standard supplement to classroom instruction, it is important to consider how the ITS compares with other types of classroom instruction in increasing student achievement, whether in terms of the use of an on-screen agent or the way that the ITS provides feedback.

Conclusions

ITS in comparison with other methods of instruction. Although the stigma associated with artificial intelligence encourages the idea that human-interaction is better for learners, all studies cited have shown that ITS has a positive impact on student learning. Holdich and Chung (2003), Kulik and Fletcher (2016), Ma, t al.(2014), Mostow et al. (2013), and Steenbergen-Hu and Cooper (2014) all found that ITS shows a positive increase in student achievement except in the case of small-group instruction or individualized human tutoring. Xin, et al. (2017) found that small-group instruction and ITS instruction were equivalent in impact on student achievement, but the ITS instruction resulted in student achievement growth at a faster rate. The reality of current school classrooms is that small group instruction of fewer than eight students and individualized tutoring are often not possible. In these situations, it is proven that ITS can have a positive impact on student achievement, especially in the case of instruction such a writing (Holdich and Chung, 2013) and guided oral reading (Mostow et al., 2013) when one-on-one instruction is necessary for the greatest impact on student achievement. Ultimately, I believe this is due to the fact that ITS are able to provide individualized feedback and extended practice that traditional classroom instruction is not, as was found by both Martin et al. (2013) and Sosa et al. (2011). In classrooms with excess of 20 students, a single teacher cannot provide instruction to this level of personalization or practice that will receive attentive feedback in the way that ITS are able to provide.

The on-screen tutor. Does an on-screen tutor, or animated pedagogical agent, increase student achievement? I believe, ultimately, that this depends on the style of instruction the student is receiving from the ITS. While studies have shown that the on-screen tutor is not beneficial and can even be distracting in linear learning environments such as students learning from presentation-style instruction (Frechette & Moreno, 2010), these linear learning environments are not typically environments that would encourage the use of the on-screen tutor. In learning environments that involve the learner receiving feedback or instruction from the ITS, the on-screen tutor can not only use gestures to enhance understanding (Cook et al., 2016) but can improve the learners experience of the ITS (Kim, 2016; McLaren, et al., 2010; Moreno & Flowerday, 2006).

Although no studies resulted in significant findings about the impact of the on-screen tutor's gender or ethnicity on student achievement overall, both Kim (2016) and Moreno and Flowerday (2006) saw signs that these qualities have a larger impact on minority students. These findings hint that the on-screen tutor could have some empowering effects on minority students and could encourage their learning.

I also believe that findings from Mayer et al. (2005) and McLaren et al. (2010) about learner sensitivity to on-screen tutor politeness hints that these tutors could be important in ITS systems that require learners to use and apply the tutor's feedback. The choice of all students to have an on-screen tutor per Cook et al. (2016) shows that students may be more receptive to feedback from a human-like tutor than from a dialogue box.

Ultimately, I believe that an on-screen tutor could have a positive impact on student achievement if used intentionally within an ITS system to enhance instruction or to provide students with human-like feedback. When used in this way, the tutor could also have positive social-emotional effects on the learner.

ITS personalization. Studies have shown for years that personalized learning has a bigger impact on student achievement than standardized learning, so much so that it has become common knowledge within the field of education. All three studies indicate that this remains true within ITS. Whether in regards to using ontology to correctly provide feedback to students (Icoz et al., 2015), personalizing questions based on student interest (Fanscali & Ritter, 2014; Walkington & Bernacki, 2019), implementing an ITS in a game-based format (Tanner & McNamara, 2013) or using student information to further personalize feedback (Crossley et al., 2016), all forms of personalization had a positive impact on student achievement, whether in final achievement scores or simply resulting in better student engagement. Although the latter seems less significant, engaging students in their learning is often the first step in obtaining higher student achievement.

Recommendations

Intelligent Tutoring Systems have been shown to have some positive effects on student achievement. Although the use of ITS would be more challenging in a classroom that does not have easily accessible computing devices, ITS should be used when possible to supplement classroom instruction

Using ITS in schools. ITS should be used to supplement instruction in classrooms. ITS allows students to receive one-on-one instruction and feedback in a way that is impossible for teachers to achieve in a traditional classroom setting (Holdich & Chung, 2003; Kulik & Fletcher, 2016; Ma, et al., 2014; Mostow et al., 2013; Steenbergen-Hu & Cooper, 2014) . ITS tools such as the on-screen tutor should be taken into account when choosing the ITS to use as it may encourage students to be more receptive to the ITS feedback, especially when engaging in new learning (McLaren et al.,

2010) and may improve their experience with the ITS (Cook, et al., 2016). The level of personalization available within the ITS should also be taken into consideration, as the higher the level of personalization, the more effective the ITS will be in increasing student achievement (Crossley, et al., 2016; Fanscali & Ritter, 2014; Icoz et al., 2015; Jackson & McNamara, 2013; Walkington & Bernacki, 2019.)

Due to the nature of ITS requiring learners to comprehend and apply the ITS feedback, younger students should not use the ITS without some teacher monitoring and intervention as necessary. The use of an ITS with an on-screen tutor may be especially beneficial for these younger students as it allows the feedback to be more human-like.

Additionally, studies have shown that ITS are more effective when students have a more prior knowledge (Sosa et al., 2011). When used as a supplement to instruction, ITS can be used with students who are making progress in the content, while teachers provide small group or individualized instruction to students who are having more difficulty mastering the content, as small group and individualized instruction have been proven to be the most effective form of instruction in increasing student achievement (Crossley, et al., 2016; Fanscali & Ritter, 2014; Icoz et al., 2015; Jackson & McNamara, 2013; Walkington & Bernacki, 2019.)

Although ITS will not likely be a magic solution to provide students with the one-on-one instruction and feedback needed to make significant growth, the ITS provides instruction that is beneficial to supplement teacher-directed classroom instruction.

Further Research

In future studies, the comparison of ITS with teacher-led instruction should be further examined, especially as ITS becomes more nuanced and personalized. It would also be beneficial to consider how an on-screen tutor that represented an adult version of the learner using the ITS could positively impact learners, especially those of minority students. It has been discussed within the educational world that minority students often do not feel empowered when their teachers do not look like them. It is possible that this same phenomenon would impact student learning within an ITS. Using an adult version of the learner could provide more information about this hypothesis.

It may also be beneficial to study whether ITS tutors that represent the learner are helpful when they are assigned to the learner based on entered data instead of giving the student a choice. Some studies indicated that student choice in regards to the on-screen tutor caused the student to become somewhat distracted by the tutor (Moreno & Flowerday, 2006). Representational tutors may be more effective if the student is not involved in choosing the tutor they felt represented them.

Finally, it would be beneficial to consider further impact of ontology to better personalize student learning. Ontology seems to be an effective method for accounting for learning pathways and connections in providing feedback to students and guiding students to increase their learning. Additionally, further studies on whether student engagement is important in regards to ITS personalization, or whether engagement can be overlooked in favor of gains in student achievement. While student engagement is evident when student interest is taken into account, a high-level of interest can also provide the opposite effect in that students become too focused on factors such as game-based aspects and are distracted from the learning.

Ultimately, research on ITS is still in its beginning stages. Further research on how to make such systems most effective is forthcoming as artificial intelligence becomes more advanced.

References

Baker, R. S., Corbett, A. T., Koedinger, K. R., & Wagner, A. Z. (2004). Off-task behavior in the cognitive tutor classroom: When students "game the system.” *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 383-390). New York, NY: Association for Computing Machinery.

Cook, S. W., Friedman, H. S., Duggan, K. A., Cui, J., & Popescu, V. (2016). Hand gesture and mathematics learning: Lessons from an avatar. *Cognitive Science, 41*(2), 518-535. doi:10.1111/cogs.12344

Crossley, S. A., Allen, L. K., Snow, E. L., & McNamara, D. S. (2016). Incorporating learning characteristics into automatic essay scoring models: What individual differences and linguistic features tell us about writing quality. *Journal of Educational Data Mining*, *8*(2), 1-20.

Differentiation. 2019. In Merriam-Webster.com. Retrieved March 8, 2019, from https://www.merriam-webster.com/dictionary/differentiation.

Fanscali, S. E., & Ritter, S. (2014). Context personalization, preferences, and performance in an intelligent tutoring system for middle chool mathematics. *Proceedings of the Fourth International Conference on Learning Analytics And Knowledge*, 73-77. New York, NY: Association for Computing Machinery.

Frechette, C., & Moreno, R. (2010). The roles of animated pedagogical agents’ presence and nonverbal communication in multimedia learning environments. *Journal of Media Psychology,22*(2), 61-72. doi:10.1027/1864-1105/a000009

Holdich, C., & Chung, P. (2003). A ‘computer tutor’ to assist children develop their narrative writing skills: Conferencing with HARRY. *International Journal of Human-Computer Studies, 59*(5), 631-669. doi:10.1016/s1071-5819(03)00086-7

Icoz, K., Ozdemir, E.B., Sanalan, V.A., Kaya, S., Cakar, M.A. (2015). Using students' performance to improve ontologies for intelligent e-learning system. *Educational Sciences: Theory and Practice, 15*(4), 1039-1050.

IXL Learning. (2019). IXL. Retrieved from https://www.ixl.com

Jackson, G. T., & McNamara, D. S. (2013). Motivation and performance in a game-based intelligent tutoring system. *Journal of Educational Psychology, 105*(4), 1036-1049. doi:10.1037/a0032580

Johnson, A. M., McCarthy, K. S., Kopp, K. J., Perret, C. A., & McNamara, D. S. (2015). Adaptive reading and writing instruction in iSTART and W-Pal. In V. Rus & Z.

Markov (Eds.), *Proceedings of the Thirtieth International Florida Artificial Intelligence Research Society Conference* (pp. 561-566).

Kazi, H., Haddawy, P., & Suebnukarn, S. (2010). Leveraging a domain ontology to increase the quality of feedback in an intelligent tutoring system. In V. Aleven, J. Kay, & J. Mostow (Eds.), *Intelligent Tutoring Systems: Tenth International Conference,* (pp. 75-84). Berlin, Germany: Springer-Verlag.

Kim, Y. (2016). The role of agent age and gender for middle-grade girls. *Computers in the Schools,33*(2), 59-70. doi:10.1080/07380569.2016.1143753

Kim, K.R. & Moon, N.M. (2013). Designing a social learning content management system based on learning objects. *Multimedia Tools and Applications*, *64*: 423. https://doi-org.proxy.lib.uni.edu/10.1007/s11042-012-1014-3

Ku, H., & Sullivan, H. (2000). Personalization of mathematics word problems in Taiwan. *Educational Technology Research and Development, 48*(3), 49–59.

Kulik, J. A., & Fletcher, J.D. (2016). Effectiveness of intelligent tutoring systems: A meta-analytic review. *Review of Educational Research*, *86*(1), 42–78.

Ma, W., Adesope, O. O., Nesbitt, J. C., & Liu, Q. (2014). Intelligent tutoring systems and learning outcomes: A meta-analysis. *Journal of Educational Psychology,106*(4), 901-918.

Martin, F., Klein, J. D., & Sullivan, H. (2006). The impact of instructional elements in computer-based instruction. *British Journal of Educational Technology, 38*, 623–636. doi:10.1111/j.1467-8535.2006.00670.x

Mayer, R. E., Johnson, W. 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*(1), 36-42. doi:10.1016/j.ijhcs.2005.07.001

McLaren, B. M., Deleeuw, K. E., & Mayer, R. E. (2010). Polite web-based intelligent tutors: Can they improve learning in classrooms? *Computers & Education, 56*(3), 574-584. doi:10.1016/j.compedu.2010.09.019

Moreno, R., & Flowerday, T. (2006). Students' choice of animated pedagogical agents in science learning: A test of the similarity-attraction hypothesis on gender and ethnicity. *Contemporary Educational Psychology, 31*(2), 186-207. doi:10.1016/j.cedpsych.2005.05.002

Mostow, J., Nelson-Taylor, J., & Beck, J. E. (2013). Computer-guided oral reading versus independent practice: Comparison of sustained silent reading to an automated reading tutor that listens. *Journal of Educational Computing Research,49*(2), 249-276. doi:10.2190/ec.49.2.g

Renninger, K., & Su, S. (2012). Interest and its development. In R. M. Ryan (Ed.), *Handbook of human motivation*. NY: Oxford University Press.

Renninger, K. A., & Hidi, S. (2016). *The power of interest for motivation and engagement*. New York, NY: Routledge.

Rock, M. L., Gregg, M., Ellis, E., & Gable, R. A. (2008). REACH: A framework for differentiating classroom instruction. *Preventing School Failure, 52*(2), 31-47. doi:https://doi-org.proxy.lib.uni.edu/10.3200/PSFL.52.2.31-47

SMARTeacher Inc. (2011, June). Prodigy. Retrieved from www.prodigygame.com

Sosa, G. W., Berger, D. E., Saw, A. T., & Mary, J. C. (2011). Effectiveness of computer-assisted instruction in statistics: A meta-analysis. *Review of Educational Research, 81*, 97–128. doi:10.3102/0034654310378174

Steenbergen-Hu, S., & Cooper, H. (2014). A meta-analysis of the effectiveness of Intelligent Tutoring Systems (ITS) on college students' academic learning. *Journal of Educational Psychology, 106*, 331–347. doi:10.1037/a0034752

Walkington, C. & Bernacki, M.L. (2019). Personalizing algebra to students' individual interests in an intelligent tutoring system: Moderators of impact. *International Journal of Artificial Intelligence in Education*, *23*, 58-88. https://doi-org.proxy.lib.uni.edu/10.1007/s40593-018-0168-1

Xin, Y. P., Tzur, R., Hord, C., Liu, J., Park, J. Y., & Si, L. (2016). An intelligent tutor-assisted mathematics intervention program for students with learning difficulties. *Learning Disability Quarterly, 40*(1), 4-16. doi:10.1177/0731948716648740