Journal of Applied Linguistics Studies Vol.2, No.1, 2023: 30-42
https://jals.aliabad.iau.ir
ISSN: 2820-9974

Intelligent (Language) Tutoring Systems: A Second-order Meta-analytic Review

Hossein Heidari Tabrizi*[1], Mahmoud Jafarie[1]

[1]Department of English Language, Isfahan (Khorasgan) Branch, Islamic Azad University, Isfahan, Iran

E-mail: m.jafarie@yahoo.com

* Corresponding Author's E-mail: heidaritabrizi@gmail.com

Received: 25-09-2022, Accepted: 03-11-2022

ABSTRACT

Intelligent Tutoring Systems (ITSs) are referred to as computerized learning environments that incorporated pedagogical, cognitive learning, knowledge representation theories into computational models of tutor, learner, and domain knowledge respectively in order to provide individualized instruction in diverse subject areas to help learners acquire domain-specific, cognitive and metacognitive knowledge. Since they take advantages of many technological artifacts of artificial intelligence (AI), knowledge representation models, and computational linguistics, they have been both object and means of research into AI, pedagogy, psychology, applied linguistics etc. Under the circumstances, the bulk of research findings is expectable and unavoidable. Sometimes researchers are badly in need of conducting first- or second-order meta-analytic reviews. This study first introduces the components of a typical ITS; then provides a descriptive account of the effectiveness, potential, and requirements of ITSs for foreign language teaching and learning (FLTL); and finally and most importantly, as a second-order review, synthesizes and describes the results of some previous meta-analytic reviews in reference to a few prevailing themes in the ITS research area such as the effects of the forms of instruction, subject matter variation, ITSs' pedagogical features etc. Toward the third end, exclusion and inclusion procedures were applied to select seven target reviews using key phrases scheme. The thematic meta-synthesis showed that the findings of almost all these reviews are consistent and congruent with some slight variations due to the study features. It is implied that meta-analyses like this can, by implication, inform and advance the science of ITS design, i.e., a theory of ITS design.

KEYWORDS: Artificial Intelligence; Foreign Language Teaching and Learning; Intelligent Language Tutoring System; Intelligent Tutoring System; Natural Language Processing; Natural Language Understanding

INTRODUCTION

Intelligent tutoring systems (ITSs) are referred to as computerized learning environments that incorporated pedagogical, psychological, and other cognitive learning theories into computational models (Luckin et_al., 2016; Graesser et al., 2011). ITS research is multidisciplinary encompassing AI, pedagogy, psychology, and other related disciplines (e.g., Craig et al., 2004; Graesser et_al., 2012; Luckin et_al., 2016; Steenbergen-Hu & Cooper, 2014). The instructional efficacy of ITS has been evaluated at all levels in a wide range of knowledge domains including physics (Albacete & VanLehn, 2000), algebra (Koedinger, Anderson, Hadley, & Mark, 1997), law (Pinkwart, Ashley, Lynch, & Aleven, 2009), medical physiology (Woo et al., 2006), language learning (Tsiriga & Virvou, 2004), reading comprehension (Mostow et al., 2002), and meta-cognitive skills (Mitrovic, 2003).

Since 1990s the most important ITS meta-analytic reviews have compared the learning effects of different ITS systems, the teaching methods, or the learning conditions (Kulik & Fletcher, 2015; VanLehn, 2011; Steenbergen-Hu & Cooper, 2014; Ma et al., 2014). Moreover, there were some systematic reviews that focused on ITS composition, current research foci, and current trends (Baker et al., 2008). Other reviews focused on technology/evaluation methods in different ITSs (Desmarais & Baker, 2012). In a recent review, Mousavinasab et al. (2021) concluded that the most frequent AI techniques applied to ITSs were condition action rule-based, Bayesian networks, and data mining.

Still other reviews have given suggestions on specific teaching mechanisms, such as strategic decision making based on student emotions (Sharma et al., 2014). The aim of this review is first to introduce the main components of a typical ITS. Then, it provides a descriptive account of the effectiveness, potential, and requirements of ITSs for FLTL. Finally, the review per se synthesizes and describes the results of previous meta-analytic reviews in relation to some prevailing themes in the research area of ITSs such as the instructional intervention conditions, ITSs' pedagogical features etc.

LITERATURE REVIEW

THEORETICAL UNDERPINNINGS OF ITS

Scholars in psychology provided theoretical foundations that were of great significance to the ITS research (Arroyo et al., 2009; D'Mello et al., 2007). For example, Kort et al. (2002) presented a comprehensive four-quadrant framework that explicitly connected learning to affective states. Moreover, Anderson (1980, 1983) proposed an adaptive control of thought (ACT) cognitive theory which became the theoretical basis for the popular Cognitive Tutor ITS system. Besides, they were Corbett and Anderson (1994) who first proposed a knowledge tracing (KT) framework based on a hidden Markov model to determine varying cognitive states during the acquisition of knowledge through analyzing student data as well as predicting future performances.

A TYPICAL ARCHITECTURE OF ITS

The requirements of the information age and the changing needs of the new generations pushed education to use new technologies and approaches when designing learner-centered environments. The first-generation computer tutoring systems have been called CAI tutors (computer-assisted instruction tutors); the second-generation tutors are usually named intelligent tutoring systems (VanLehn, 2011).

According to Ma et al. (2014), an ITS is a computer system that for each student:

1. Performs tutoring functions by (a) presenting information to be learned, (b) asking questions or assigning learning tasks, (c) providing feedback or hints, (d) answering questions posed by students, or (e) providing prompts to provoke cognitive, metacognitive or motivational change,
2. By computing inferences from student responses constructs either a persistent multidimensional model of the student's psychological states (such as subject matter knowledge, learning strategies, motivations, or emotions) or locates the current cognitive state of the student in a multi-dimensional domain model,
3. Uses the student modeling functions identified in point 2 to adapt one or more of the tutoring functions identified in point 1.

To perform the tutoring functions as stated in point 1 above, ITSs are assumed to have a four-component conceptual structure that has been sustainable, even as ITSs themselves varied significantly in their design. The four "generally accepted" (Sottilare et al., 2013, p. ii) conceptual components of ITSs are as follows:

1. *An interface* that communicates with the learner by presenting and receiving information. The interface, often confined to the subject domain (e.g., algebra), specifies the moves that the learner can make in seeking information, solving problems or responding to questions.

2. *A domain model* that represents the knowledge the student is intended to learn. This model is a set of production rules, logical propositions, natural language statements, or any appropriate knowledge representation format.

3. *A student model* that represents relevant aspects of the student's knowledge determined by the student's responses to questions or other interactions with the interface. While the student model may be an overlay or subset of the domain model, in some ITS it represents common misconceptions or other "bugs" in the student's knowledge.

4. *A tutor model* that represents instructional strategies such as offering a hint when the student is unable to generate a correct response or assigning a problem that requires knowledge only slightly beyond the current student model.

In addition to the above conceptual components, the ITSs with natural language support need some kind of natural language processing (NLP) tool. It is, as a computational tool, used to learn, understand, and generate human language content. NLP is also employed to convert information stored in natural language to a machine-understandable format. NLP-enabled systems differ in the employed techniques, are developed for different purposes, and can vary in focus to link natural language inputs and outputs with their domain models. NLU, as an offshoot of NLP, addresses the understanding of the user's utterances (Paladines & Ramirez, 2020). For any intelligent tutoring system to attempt a deep understanding of students' natural language explanations, it needs to choose between three main existing basic approaches: symbolic, statistical, and hybrid.

The symbolic approach: This approach is the traditional approach for mapping NL to a knowledge representation language. It is based on linguistic or lexicographical knowledge. This knowledge is elicited and specified by a language expert in form of a lexicon and a grammar that can later be used to parse sentence/utterance strings as the user inputs (Jordan et al., 2004; Paladines & Ramirez, 2020). Many practical and robust sentence-level syntactic parsers exist for which there are wide coverage NL lexicons and grammars (Abney, 1996). However, syntactic analysis can only be licensed in relation to syntactic aspects of lexical semantics (Levin & Pinker, 1992). For instance, the similarity of "I baked a cake for her" and "I baked her a cake" is recognized but the similarity of this pair to "I made her a cake" is not. It needs a sort of licensing that is typical of semantic analysis. Now there is no general solution at this level since semantic analysis has to do with the realm of cognition and mental representations and is to be designed in relation to the interested domain (Jackendoff, 1983).

The statistical approach: This more recent approach to processing NL exploits a corpus to train a wide variety of statistical methods for analyzing language. It is popular because it will easily work if there already is a representative corpus. The most useful method for ITSs has been text classification in which a subtext is tagged as being a member of a particular class of interest and uses just the words in that tagged subtext for training a classifier. This kind of classification is known as a *bag of words* method because it does not take care of the meaning that a sentence word order communicates. With the help of this approach, the text can be linked to its representation by searching for a hand-generated propositional representation for the exemplar text of the class identified at run-time. Most ITSs include a sentence segmenter that attempts to break up complex sentences prior to sentence understanding by any of the methods (Jordan et al., 2004).

The hybrid approach: This approach results from a combination of the other two approaches pursuing the goal of complementing each other. It is the one in which rich, symbolically produced syntactic features can be used to supplement the training of a text classifier (instead of the syntactic features obtained via statistical parsing methods). As for text classification, the classes are generally specified via a prototypical subtext of the class so the desired propositional representation must still be obtained via a look-up according to the class identified at run-time.

THE EFFECTIVENESS, POTENTIAL, AND REQUIREMENTS OF ITSS FOR FLTL

Action research is intended to contribute to the process of schooling as a whole through the agency of the teacher and the co-operation of the learners: to design quality courses; to better develop, evaluate, and adapt instructional materials including textbooks; and to improve or integrate the processes of teaching and assessing. Moreover, personalized instruction, as it refers to tailoring instruction to each learner's educational needs and preferences, has been a dominant approach to education in general and foreign language learning and teaching in particular. How could it be possible for a teacher to do all these jobs spontaneously and simultaneously with regard to the prevailing constraints such as time, budget, curricular, societal, and expertise? A teacher does never afford, or dare attempt, such a massive risky undertaking! The first and foremost possibility of addressing the issue that comes to our mind is to give responsibility to an alternative agent competent and expert enough to take care of this. And it is not but an intelligent tutoring system.

A part of the advantage of ITSs over traditional classroom instruction can be related to the features ITSs share with other forms of computer-based instruction (CBI). They are as follows: greater interactivity and adaptation, greater immediacy of feedback, response-specific feedback, greater cognitive engagement, more opportunity for practice and feedback, increased learner control, and individualized task selection. It is presumably the multidimensional student modeling that enables ITS to outperform non-ITS CBI on each of its advantages mentioned in the above paragraph. An ITS that also models domain knowledge as production rules can do the task selection by qualifying not only each task as a set of production rules required to complete it but also each student as a set of production rules that most need to be practiced, and then finding the best match. This multidimensional matching is likely to be more effective than unidimensionally matching student ability to task difficulty (Ma et al., 2014).

Assuredly, a significant advantage of ITS or more precisely ILTS is error-specific learner feedback, that is, an intelligent computer feedback which explains the type of an error. Research shows that is more effective that other traditional feedbacks (Yang & Akahori, 1999). An ILTS should provide meaningful and interactive vocabulary and grammar practice for EFL learners. It should analyze the learners' sentences and detect the grammatical and other errors. The main function of the feedback module is to correlate the detailed output of linguistic analysis with an error-specific feedback message. And most importantly, the feedback is individualized using an adaptive student model (SM). The SM monitors a learner's performance over time for different grammatical constructions. This record is used to gear feedback messages to learner proficiency within a framework of guided discovery learning. A novice will receive the most explicit jargon-free feedback whereas that of the advanced learner will be just hint at error. The SM does this by keeping score for each grammatical structure. The score for each node will increase or decrease depending on the learner input. There is a master file which specifies the amount by which the scores of each node are adjusted and is perhaps weighted to reflect such different pedagogical purposes as importance of an error, significance of an exercise, etc. Therefore, feedback messages are tuned to the current learner level at each grammatical structure. Learners start out at the intermediate level. If they were below that level, the system nodes for each grammatical construct would adjust quickly owing to the detailed scoring (Toole & Heift, 2002).

SECOND-ORDER META-ANALYTIC REVIEW OF ITS RESEACH

Prior major meta-analytic reviews of ITSs are as follows. First, VanLehn (2011) found that ITSs and human tutoring learning effect sizes are roughly the same (0.76, 0.79). In other words, they are as effective as human tutoring. Second, Steenbergen-Hu and Cooper's (2014) review made previous answers to questions related to ITS's effectiveness on learning clearer. Accordingly, ITSs have showed their ability to outperform many learning activities or instructional methods in making easier college students' learning of divers subjects though human tutors are still more effective than them. Third, meta-analysis conducted by Ma et al. (2014) investigated how effect sizes changed with ITS type, type of learning outcome, type of treatment received by learners, whether declarative or procedural knowledge was to be learned, and other variables. Forth, Kulik and Fletcher (2015) describes that he median effect of intelligent tutoring in the 50 evaluative studies was to increase test scores 0.66 standard deviations over conventional levels depending to a great extent upon whether improvement was gauged on standardized or locally developed tests. Fifth, the review of Paladines and Ramírez (2020) revealed that the majority of ITSs are aimed at science, technology, engineering, and mathematics (STEM) domains at the higher education levels and most of the selected ITSs are of the expectations and misconceptions tailoring (EMT) type. Moreover, a great number of ITSs make use of dialog to help students learn how to solve a problem through applying laws, rules as well as levels in the Bloom's (1956) taxonomy. As to the instructional approach, the ITSs of choice help students write correct answers or explanations for deep questions; aid them in problem solving; or back a reflective dialogue encouraged by either the result of a simulation or previously provided content. Sixth, Atun's (2020) study critically reviewed the studies on the development of reading comprehension skills using ITSs. It concluded that ITSs (such as ITS for the Text Structure Strategy) are more effective than the traditional teaching methods in developing reading comprehension skills and the results were biased in favor of the students with low reading ability. Seventh, Guo's (2021) study surveyed future trends and the evolution of ITS research via scientometric methods. It found that ITS research has been growing in recent years; that education, computer science, engineering, and psychology were the main ITS research knowledge sources; and that student modeling, interactive learning environments, machine learning, and teaching/learning strategies have been the foci of most popular research.

There are many incentives and needs that necessitate the review of the ITS effectiveness on language education, in particular on FLTL. First, few existing research syntheses of the effectiveness of educational technology in language education have focused on ITS though most usually cover a broad range of technologies. They encompass computer-assisted language learning (CALL), CBI, word processor, hypermedia, information and communication technology (e-mail and text-messaging applications), simulation, and digital media. Therefore, there is a need to distinguish ITS from other types of educational technologies and conduct a separate systematic summarization of its effectiveness in relation to FLTL. Second, researchers have so far conducted some systematic reviews on ITS's effectiveness on such subject domains as algebra, basic mathematics, statistics, physics, computer sciences, economics, but few exists for FLTL. Third, we need to know whether and how ITS affects students' learning differently in different subject domains.

This review was intended to do a second-order meta-analysis of some previous reviews of research into intelligent tutoring systems (ITSs) in relation to FLTL. On the basis of recent meta-analyses of ITS effectiveness, it attempted to provide a descriptive account of the effectiveness, efficiency and potential of ITSs for specific subject matters as well as the degree of their relevance and usability to FLTL. It did also deal with the requirements, in case they are fulfilled, for ITSs to be canonicalized for the cause, i.e., FLTL. To these ends, it addresses the following research questions.

RQ1: Is ITS more effective than other forms of instruction? How effective are they?
RQ2: For what purposes, target audiences, and subject matters are ITSs developed?
RQ3: Are ITSs' effects on the learning and performance of learners different from one subject matter to another one?
RQ4: Has an ITS to be used standalone or integrated with regular classroom instruction?
RQ5: What instructional approaches and support resources do ITSs make use of?
RQ6: How have ITS knowledge sources been developed?
RQ7: How many percent of the existing ITSs are devoted to FLTL or are of general purpose, i.e. are not used for tutoring certain subject matters?

Journal of Applied Linguistics Studies Vol.2, No.1, 2023: 30-42
https://jals.aliabad.iau.ir
ISSN: 2820-9974

METHODOLOGY

RESEARCH METHOD

In this study, the second-order research method of meta-synthesis, which DeWitt-Brinks and Rhodes (1992) also referred to as meta-analysis, was used to examine, interpret, and synthesize parts of multiple studies that are mostly related to specific- and general-purpose ITSs and partly to ILTSs or FLTL ITSs. Despite slight variations, the meta-analysis process is generally carried out as follows:

- The analyst first determines the reason, aim and questions of the research,
- Then s/he conducts a literature review with keywords appropriate for the purpose of the research,
- Third, s/he determines the sources that are suitable for the purpose of the research using methods of inclusion and exclusion,
- Forth, s/he analyzes the selected studies according to appropriate themes and subthemes, and
- Finally, s/he reports and discusses the results (Bartoș & Langdridge, 2019).

DATA COLLECTION/SEARCH STRATEGY

There are different views on the number of studies to be examined in meta-analytic reviews. Weed (2005) refers to meta-synthesis as the interpretation of a few qualitative studies and did not suggest the number of studies. For this second-order meta-synthesis, meta-analytic reviews with the key phrases ‘tutoring’, ‘tutoring system’, ‘intelligent tutoring system’, ‘adaptive intelligent tutoring system’, and ‘intelligent language tutoring systems’ in the World Wide Web were identified.

INCLUSION AND EXCLUSION OF STUDIES

Selection criteria were specified in such a way that selected reviews shall be of quite relevance and potentiality to general domains and/or FLTL.

The search results that met the following ***criteria*** were included in the study, and those that did not were excluded.

- One that compares the effectiveness of ITSs with other forms of instruction.
- One that reviews the studies of ITSs treating college students.
- One that reviews the studies of LTSs in relation to features of the study.
- One that reviews the studies of ITLs in relation to the embedded technologies.

CODING STUDY CHARACTERISTICS AND THEMES

The studies were already coded by the previous reviewers. As for the intended themes, they were derived based on the following rubric:

- The effects of the forms of instruction,
- Reasons for the development of ITSs
- Subject matter variation,
- The instructional intervention conditions,
- ITSs’ pedagogical features,
- Multidisciplinary nature of ITS studies, and
- The market share of ILTSs or FLTL ITSs.

DATA ANALYSIS AND INTERPRETATION

The selected reviews for examination (Table 1.) were descriptively analyzed in the light of the pre-determined themes. The analysis used meta-synthesis methods within the scope of the research design.

Journal of Applied Linguistics Studies Vol.2, No.1, 2023: 30-42
https://jals.aliabad.iau.ir
ISSN: 2820-9974

Table 1.
The Included Meta-Analytic Reviews

No	Authors	Year	Title
1.	VanLehn, K.	2011	The Relative Effectiveness of Human Tutoring, Intelligent Tutoring Systems, and Other Tutoring Systems
2.	Steenbergen-Hu, S. & Cooper, H.	2014	A Meta-Analysis of the Effectiveness of Intelligent Tutoring Systems on College Students' Academic Learning
3.	Ma, W., Adesope, O. O., Nesbit, J. C., & Liu, Q.	2014	Intelligent Tutoring Systems and Learning Outcomes: A Meta-Analysis
4.	Kulik, J. A. & Fletcher J. D.	2015	Effectiveness of Intelligent Tutoring Systems: A Meta-Analytic Review
5.	Paladines, J. & Ramírez, J.	2020	A Systematic Literature Review of Intelligent Tutoring Systems With Dialogue in Natural Language
6.	Atun, H.	2020	Intelligent tutoring systems (ITS) to improve reading comprehension: A systematic review.
7.	Guo, L. et al.	2021	Evolution and Trends in Intelligent Tutoring Systems Research: A Multidisciplinary And Scientometric View

RESULTS AND DISCUSSION

The descriptive account is to answer the following research questions.

RQ1: Is ITS more effective than other forms of instruction? How effective they are?

Theme one: effects of the forms of instruction

First, the meta-analysis of Steenbergen-Hu and Cooper (2014) examined the overall effectiveness of ITS on college students' learning, measured with adjusted effect sizes. Under a fixed-effect model, the average effect size (*ES*) was g = .37, 95% confidence interval (CI) [.24, .50], p = .000, and was significantly different from zero. Under a random-effects model, the average effect size was also g = .37, 95% CI [.21, .53], p = .000, and was significantly different from zero. The effect sizes appeared to be homogeneous, $Q_t(25) = 35.47$, p = .080, f = 29.51.

Second, Ma et al. (2014) hypothesized that a portion of the advantage of ITS over traditional classroom instruction and learning activities with printed materials can be attributed to the features ITS share with other kinds of CBI. Likewise, others have generally justified the CBI advantage as resulting from greater interactivity and adaptation than is available in other instruction modes, and specifically attributed the effectiveness of CBI to greater immediacy of feedback, feedback that is more response-specific, greater cognitive engagement, more opportunity for practice and feedback, increased learner control, and individualized task selection. They also hypothesized that multidimensional student modeling enables ITS to outperform non-ITS instruction modes on each of its advantages mentioned above. According to them, the effect sizes, generated by Comprehensive Meta-Analysis software, are mainly spanned between –0.25 and 0.75 standard deviations after adjusting the outliers indicated that in most studies the ITS comparison groups outperformed their respective control groups—a positive effect size indicating that students who used ITS outperformed those who experienced other modes of instruction and vice versa.

Third, Kulik and Fletcher (2015) conducted two analyses: primary and supplementary analyses. The primary analysis used the estimator of effect size called Glass's *ES*, evaluation study as the unit of analysis, and unweighted means to represent combined effects. The latter used the estimator of effect size called Hedges's *g*, both evaluation study and evaluation finding as units of analysis, and both weighted means and unweighted means to represent overall effects.

Based on their primary analysis, students who received intelligent tutoring outperformed control students on posttests in 46 (or 92%) of the fifty studies. In 39 (or 78%) of the 50 studies, tutoring gains were larger than 0.25 standard deviations, or large enough to be considered of substantive importance by the standards of the What Works Clearinghouse (Institute of Education Sciences, U.S. Department of Education, What Works Clearinghouse, 2013). Therefore, a good number of studies found ITS impacts that were both positive and large enough to be significant for instruction. An increase in test scores of 0.66 standard deviations over normal levels is equivalent to an increase from the 50th to the 75th percentile. By Cohen's (1988) standards, the average *ES* for intelligent tutoring is moderate to large where ES is the Glass's estimator of effect size.

According to their supplementary analysis, results were affected slightly by this change in unit of analysis. For instance, in the dataset of 63 independent comparisons, the median ES is 0.63. In 58 (or 92%) of the 63 comparisons, the ITS group scored higher than the control group; and in 49 (or 78%) of the comparisons, the improvement due to ITS use was substantively important, or more than 0.25 standard deviations. It is important to note that ITSs' effects were very large in some studies and near zero in others. Therefore, in order to determine whether study features (such as test type, sample size, grade level, etc.) were related to the variation in results, they carried out a series of univariate analyses of variance (ANOVAs) with study feature as independent variable and size of effect as dependent variable. Each set of ANOVAs showed that test type was the study feature most strongly related to size of effect. Finally, Atun (2020) observed that in all the studies conducted in the Language Arts course the effect of ITS on reading comprehension skill is positive. In other words, ITSs are effective in increasing students' reading comprehension skill regardless of student differences.

RQ2: For what purposes, target audiences, and subject matters are ITSs developed?

Theme two: reasons for ITSs development, target students, and subject contents

Tabular data included in the relevant studies showed that ITSs are engineered for reasons such as adapting instruction to certain criteria, using technology more effectively and efficiently in education, making teaching materials and environment more interesting, and providing easy-to-use teaching materials. With regard to end users the same data indicated that these systems are designed for students at all levels of education from primary school to graduate studies. Moreover, as far as subject contents are concerned, the data showed that more than a third (38%) of the studies reported the use of ITSs in the field of Information Technologies, followed by Mathematics at 22%, while miscellaneous fields represented 10% or fewer of the studies.

RQ3: Are ITSs' effects on the learning and performance of learners different from one subject matter to another one?

Theme three: subject matter variation

Steenbergen-Hu and Cooper (2014) wanted not only to know the ITSs' effectiveness overall but also to know whether ITS affect students' learning differently in different subject domains. To examine it, they grouped the studies by subject matter. Finally, they found that ITS's effectiveness did not differ significantly depending on which subjects ITSs were used for under either a fixed-effect model, Qb = 4.01, p = .548, or a random-effects model, Q_b = 2.78, p = .734, where Q_b denotes the heterogeneity status between all categories of a particular variable.

RQ4: Has an ITS to be used standalone or integrated with regular classroom instruction?

Theme four: the instructional intervention conditions

According to Steenbergen-Hu and Cooper (2014), studies vary in the way they use ITSs as interventions. They classified the intervention conditions into five categories, each one demonstrating different manners or levels of ITS's involvement in the intervention. In majority of the studies, students used ITS as the main or only treatment to learn subject matter; they called this intervention condition as principal instruction. In the second major group of studies, ITS was integrated into classroom instruction as it played an important part. They called it ITS-integrated class instruction. In other studies, ITS was used to supplement classroom instruction for additional learning after regular classes called it ITS-supplemented class instruction. In the forth intervention condition they labeled it ITS-assisted activities, students used ITS for laboratory or exercises that usually took place during class time. In fifth group of studies, students used ITS to do after-school homework assignment. They called it ITS-assisted homework. Such classification enabled them to explore whether ITS's effectiveness differs by the way and the degrees of its uses. They termed these five intervention conditions holistically as ITS-assisted learning. To find out whether ITSs' effectiveness differed depending on how ITSs were involved in the interventions, they grouped the studies by intervention conditions. In conclusion, they found that the average effect sizes did not differ significantly depending on how ITSs were used in the intervention under either a fixed-effect model, $Q_b = 8.52$, $p = .074$, or a random-effects model, $Q_b = 5.95$, $p = .203$.

RQ5: What instructional approaches and support resources do ITSs make use of?

Theme five: pedagogical features

This question was asked to let us know the pedagogical features of the ITSs. There are three well-established relevant features related to the tutoring process: the instructional approach, the support resources, and the input text format (Paladines & Ramirez, 2020).

The instructional approach refers to the strategy employed by the ITS to make sure that the students learn the target matter and reach their learning objectives. It is generally inspired by the theory of Socratic and constructivist learning. Paladines and Ramirez (2020) developed their own classification using four categories of approaches: generation of explanations to justify solution, support for problem solving, clarify and direct procedures, and ask questions-answer. The ITS, in the first approach, empowers the student so as to actively elaborate explanations and justifications of a previous student input (e.g., a previous student's answer, a student prediction related to a simulation) in a mixed-initiative turn-based dialog. In each turn, the system compares the student's answer with the expectations and misconceptions/errors that are already determined for each question. To address misconceptions, the ITS has a set of anticipated incorrect answers (bugs) and their relevant remedies. Paladines and Ramirez (2020) found that most of the ITSs implement this approach. The ITSs which implement the second approach can offer support for problem solving, either on demand where they detect a student's error/misconception, or when the solution is not sufficiently complete. The third instructional approach, i.e. the clarification and direction of the procedures, is a sub-approach of the previous approach. However, Paladines and Ramirez (2020) considered it a separate approach to highlight that in ITSs implementing this approach, learners need scaffolding to carry out a predetermined procedure or task in a step-wise manner. In contrast, the ITSs of the previous approach are intended to support the resolution of problems whose solution cannot be built step by step following a recipe. They suggested that this difference plays a transparent role in how the ITSs of these two approaches direct students throughout the problem-solving. Therefore, for them, it makes sense to present these two approaches separately. The forth approach, ask questions-answer, is the simplest to implement since the two ITSs of this group, Abdullah and Lana, present questions based only on the pieces of information provided to the student prior to the dialog and then wait for short answers. In case of a mistake on the part of the student, these two ITSs react in different ways. Abdullah gives another opportunity to the student to correct his/her answer. If the student fails again to provide a correct answer, Abdullah will show it to the student and proceeds with the next question. Unlike Abdullah, Lana directly explains the right answer to the student and continues with the next question.

Support resources, as a whole, is the second pedagogical feature on which the instructional approach relies. They refer to the types of resources with which each ITS supports its teaching process either during the dialogue, before or after. Either case, these resources serve to encourage the dialogue to help students build utterances/solutions, or to describe concepts to be addressed in the conversation. This feature can take some of the following eight values: animated agent, audio and video, conceptual maps, images, option menu, simulation, table and virtual slides. The results showed that animated agents and images are widely used resources by ITSs because they facilitate the learning process and can easily be incorporated into the system.

The input text format is the third pedagogical feature through which the ITSs assess the knowledge acquired by the students. Paladines and Ramirez (2020) identified two input text formats: long and short answer. While a long answer or short essay is typically composed of at least 2 sentences and is the form of evaluation of most ITSs, short answers normally consist of at most two short sentences. It is important to note that some ITSs elicit both short and long answers, the former through dialogue, and the latter, i.e. the construction of explanations, by means of option menu.

RQ6: How have ITS knowledge sources been developed?
Theme six: Multidisciplinary characteristics based on WoS category
Guo et al. (2021) let the journal categories for the ITS publications be generated from the WoS classification tag "WoS Categories" and run the CiteSpace software category function to detect more detailed information. It showed that the ideas for ITSs were rooted in the education discipline. In the 1960s, pioneering education researcher developed a primitive ITS application. Owing to some constraints, ITSs research laid dormant for a while. But, in the 1980s, a group of scholars incorporated computer technology theories and methods into ITS research, and in 1986, the psychological theories and methods were introduced to contribute to the study of evaluation. Since computer science was the main ITS function backbone and the theoretical basis for ITSs was education and psychology, there was a need for the integration to be realized through technology. Thus, the relevant computer science studies were concentrated on specialized skills computer science and the application of psychological or educational frameworks. Some scholars proposed methods to take advantage of semantic web-based approaches when representing ITS pedagogical rules, and some others applied Cognitive Diagnosis models, which were an evaluation system based on cognitive psychology, statistics, and computer science, to an ITS to model student answer data. It gave rise to a correlation between test questions and knowledge structures that enabled the diagnosis of student cognitive states and the quantitative investigation of student differences and cognitive levels.

RQ7: How many percent of the existing ITSs are devoted to FLTL or are of general purpose, i.e. are not used for tutoring certain subject matters?
Theme seven: the market share of ILITS or FLTL ITS
In this review, an inventory of 120 tutoring systems was made. It consists of ITSs both general and specific in domain as well as ICALLs. A frequency and percentage analysis showed that about a forth (20%) of the systems serve as means of learning and teaching of languages, in particular FLTL. Seven systems (5.8%) were devoted to general purposes, i.e. they were designed to teach almost every subject matter at every level. Every one of the remaining was devoted to a specific subject matter.

CONCLUSION

The current review first provided an introduction of the conceptual architecture of a typical ITS. An ITS is identified to be made at least from four different models: interface, domain, student, and tutoring, though other components such as feedback in case of intelligent language tutoring system (ILTS) may be added to it. It also takes advantage of a critical embedded element, that is, natural language processing (NLP). NLP has to do with using computational techniques to learn, understand, and produce human language content. Natural Language Understanding (NLU), as an offshoot of NLP, takes care of deep understanding the user's utterances/explanations. For any ITS to attempt a deep understanding of students' natural language explanations, it needs to choose between three main existing basic approaches: symbolic, statistical, and hybrid.

As far as the effectiveness, potential, and requirements of ITSs for FLTL are concerned, a significant advantage of ITS or more precisely ILTS certainly is error-specific learner feedback, that is, an intelligent computer feedback which explains the type of an error. Research shows that is more effective that other traditional feedbacks. Most importantly, the thematic meta-synthesis showed that the findings of almost all the ITSs' meta-analytic reviews are consistent and congruent with some slight variations due to the study features. For example, as regards the effects of different forms of instruction, all concluded that ITS treatment makes moderately significance difference comparable to other forms of teaching. Finally, a frequency and percentage analysis showed that about a forth (20%) of the ITSs serve as means of learning and teaching of languages, in particular FLTL. Application of meta-analyses like this to the products of ITS research, development, and evaluation can, by implication, inform the science of ITS design. In other words, they contribute to the advancement of a theory of ITS design.

REFERENCES

Abney, S. (1996). Partial parsing via finite-state cascades. *Journal of Natural Language Engineering, 2***,** 337–344. https://doi.org/10.1017/S1351324997001599

Albacete, P. L., & VanLehn, K. A. (2000). Evaluating the effectiveness of a cognitive tutor for fundamental physics concepts. In L. R. Gleitman & A. K. Joshi (Eds.), *Proceedings of the 22nd Annual Meeting of the Cognitive Science Society* (pp. 25–30). Mahwah, NJ: Erlbaum.

Anderson, J. R. (1980). *Cognitive psychology and its implications*. San Francisco, CA: Freeman.

Anderson, J. R. (1983). *The architecture of cognition*. Cambridge, Mass: Harvard University Press.

Arroyo, I., Cooper, D. G., Burleson, W., Woolf, B. P., Muldner, K., & Christopherson, R. (2009). *Emotion sensors go to school*. IOS Press.

Atun, H. (2020). Intelligent tutoring systems (ITS) to improve reading comprehension: A systematic review. *Journal of Teacher Education and Lifelong Learning, 2*(2), 77-89.

Baker, R. S. J. D., Corbett, A. T. & Aleven, V. (2008). More accurate student modeling through contextual estimation of slip and guess probabilities in bayesian knowledge tracing. In B. P., Woolf, E., Aïmeur, R., Nkambou, S., Lajoie (Eds.), *Intelligent Tutoring Systems. ITS 2008. Lecture Notes in Computer Science*, vol 5091. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-69132-7_44

Bartoș, S. E., & Langdridge, D. (2019). LGBQ resilience: A thematic meta-synthesis of qualitative research. *Psychology & Sexuality*,*10*(3), 234-247. https://doi.org/10.1080/19419899.2019.1596973

Bloom, B. S. (Ed.). (1956). *Taxonomy of educational objectives: The classification of educational goals, by a committee of college and university examiners*. New York: Longmans.

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

Corbett, A. T., & Anderson, J. R. (1994). Knowledge tracing: Modeling the acquisition of procedural knowledge. *User modeling and user-adapted interaction, 4*(4), 253–278.

Craig, S., Graesser, A., Sullins, J., & Gholson, B. (2004). Affect and learning: An exploratory look into the role of affect in learning with AutoTutor. *Journal of Educational Media, 29*(3), 241–250. https://doi.org/10.1080/1358165042000283101

D'Mello, S., Graesser, A., & Picard, R. W. (2007). Toward an affect-sensitive AutoTutor. *IEEE Intelligent Systems, 22*(4), 53–61. doi: 10.1109/MIS.2007.79

Desmarais, M. C., & d Baker, R. S. (2012). A review of recent advances in learner and skill modeling in intelligent learning environments. *User Modeling and User-Adapted Interaction, 22*(1), 9–38. doi:10.1007/s11257-011-9106-8

DeWitt-Brinks, D. & Rhodes, S. C. (1992). *Listening instruction: A qualitative meta-analysis of twenty-four selected studies*. (rep. No. clearinghouse: CS507954). US: Michigan.

Graesser, A. C., & D'Mello, S. (2012). Emotions during the learning of difficult material. *Psychology of Learning and Motivation, 57*, 183–225. doi:10.1016/B978-0-12-394293-7.00005-4

Graesser, A. C., Conley, M. W., & Olney, A. M. (2011). Intelligent tutoring systems. In S. Graham & K. Harris (Eds.), *APA educational psychology handbook: Vol. 3. Applications to learning and teaching.* American Psychological Association.

Guo, L., Wang, D., Gu, F., Li, Y., Wang, Y., & Zhou, R. (2021). Evolution and trends in intelligent tutoring systems research: A multidisciplinary and scientometric view. *Asia Pacific Education Review.* https://doi.org/10.1007/s12564-021-09697-7

Jackendoff, R. (1983). *Semantics and Cognition*. Current Studies in Linguistics Series. The MIT Press.

Jordan, P. W., Makatchev, M., & VanLehn, K. (2004). Combining competing language understanding approaches in an intelligent tutoring system. In J.C. Lester et al. (Eds.), *ITS 2004, LNCS 3220* (pp. 346–357), Springer-Verlag, Berlin: Heidelberg.

Koedinger, K. R., Anderson, J. R., Hadley, W. H., & Mark, M. (1997). Intelligent tutoring goes to school in the big city. *International Journal of Artificial Intelligence in Education, 8*, 30–43.

Kort, B., Reilly, R. & Picard, R.W. (2002). An affective model of interplay between emotions and learning: reengineering educational pedagogy-building a learning companion. *Proceedings IEEE International Conference on Advanced Learning Technologies*. IEEE. doi: 10.1109/ICALT.2001.943850

Kulik, J. A. & Fletcher, J. D. (2015). Effectiveness of intelligent tutoring systems: A meta-analytic review. *Review of Educational Research, 86*(1), 4278. doi:10.3102/0034654315581420

Kürşat-Erümit, A. & Çetin, I. (2020). Design framework of adaptive intelligent tutoring systems. *Education and Information Technologies*. https://doi.org/10.1007/s10639-020-10182-8

Levin, B. & Pinker, S. (1992). *Lexical and conceptual semantics*. Oxford: Blackwell Publishers.

Luckin, R., Holmes, W., Griffiths, M., & Forcier, L. B. (2016). *Intelligence unleashed. An argument for AI in education*. London: Pearson.

Ma, W., Adesope, O. O., Nesbit, J. C., & Liu, Q. (2014). Intelligent tutoring systems and learning outcomes: A meta-analysis. *Journal of Educational Psychology, 106*(4), 901–918. https://doi.org/10.1037/a0037123

Mitrovic, A. (2003). An intelligent SQL tutor on the web. *International Journal of Artificial Intelligence in Education*, *13*(2– 4), 171–195.

Mostow, J., Aist, G., Bey, J., Burkhead, P., Cuneo, A., Junker, B., . . . Wilson, S. (2002). Independent versus computer-assisted reading: Equal-time comparison of sustained silent reading to an automated reading tutor that listens. Paper presented at *the 9th Annual Meeting of the Society for the Scientific Study of Reading*, Chicago, IL.

Mousavinasab, E., Zarifsanaiey, N., Niakan Kalhori, S. R., Rakhshan, M., Keikha, L., & Ghazi Saeedi, M. (2021). Intelligent tutoring systems: A systematic review of characteristics, applications, and evaluation methods. *Interactive Learning Environments, 29*(1), 142–163. https://doi.org/10.1080/10494820.2018.1558257

Palandines, J. & Ramirez, J. (2020). A systematic literature review of intelligent tutoring systems with dialogue in natural language. *IEEEAccess, 8*, 164246- 164267. doi: 10.1109/ACCESS.2020.3021383

Pinkwart, N., Ashley, K., Lynch, C., & Aleven, V. (2009). Evaluating an Intelligent Tutoring System for making legal arguments with hypotheticals. *International Journal of Artificial Intelligence in Education, 19,* 401–424.

Sharma, S., Ghorpade, S., Sahni, A., & Saluja, N. (2014). Survey of intelligent tutoring systems: A review on the development of expert/intelligent tutoring systems, various teaching strategies and expert tutoring system design suggestions. *International Journal of Engineering Research and Technology, 3*(11), 37–42. doi: 10.17577/IJERTV3IS110028

Sottilare, R., Graesser, A., Hu, X., & Holden, H. (Eds.). (2013). *Design recommendations for Intelligent Tutoring Systems*. Orlando, FL: U.S. Army Research Laboratory.

Steenbergen-Hu, S., & Cooper, H. (2014). A meta-analysis of the effectiveness of intelligent tutoring systems on college students’ academic learning. *Journal of Educational Psychology, 106*(2), 331–347. https://doi.org/10.1037/a0034752

Toole, J. & Heift, T. (2002). The tutor assistant: An authoring tool for an intelligent language tutoring system. *Computer Assisted Language Learning, 15*(4), 373-386. doi:10.1076/call.15.4.373.8266

Tsiriga, V., & Virvou, M. (2004). Evaluating the intelligent features of a web-based intelligent computer assisted language learning system. *International Journal on Artificial Intelligence Tools, 13,* 411–425. doi:10.1142/S0218213004001600

VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist, 46*(4), 197–221. doi:10.1080/00461520.2011.611369

Weed, M. (2005). "Meta interpretation": A method for the interpretive synthesis of qualitative research. In *Forum Qualitative Sozialforschung/Forum: Qualitative Social Research (Vol. 6, No. 1)*, 1–21.

What Works Clearinghouse. (2013). *What Works Clearinghouse: Procedures and standards handbook* (Version 3.0). Retrieved from http://ies.ed.gov/ncee/wwc/pdf/reference_resources/wwc_procedures.pdf

Woo, C. W., Evens, W. M. W., Freedman, R., Glass, M., Seop Shim, L., Zhang, Y., . . . Michael, J. (2006). An Intelligent Tutoring System that generates a natural language dialogue using dynamic multi-level planning. *Artificial Intelligence in Medicine, 38,* 25–46. doi:10.1016/j.artmed.2005.10.004

Yang, J. & Akahori, K. (1999). An evaluation of Japanese CALL systems in the WWW: Comparing a freely input approach with multiple selections. *Computer Assisted Language Learning, 12*, 59-79. doi:10.1076/call.12.1.59.5717