A Chatbot Tutor Can Lessen the Gender Confidence Gap in Information Systems Learning

Cherileigh Leavitt
Brigham Young University
cherileigh.leavitt@gmail.com

Madison Corbin
Brigham Young University
maddysherwood33@gmail.com

Mark Keith
Brigham Young University
mark.keith@gmail.com

Ryan Schuetzler
Brigham Young University
ryan.schuetzler@byu.edu

Degan J. Kettles
Brigham Young University
degank@gmail.com

Greg Anderson
Brigham Young University
profganderson@byu.edu

Abstract

Women are underrepresented in the information systems discipline as well as other STEM fields. A common explanation for the significant gender gap is the difference in women's and men's self-efficacy concerning information systems (IS) tasks and roles during their secondary education. As a potential solution, this study explores how chatbot tutors impact confidence differently between 136 women and men in an introductory programming course. Our findings confirmed prior research showing that while men have greater confidence in information systems tasks, there is no difference in performance in those tasks between women and men. Next, we found that a chatbot used during learning can improve confidence of all students. Finally, and most importantly, we found that the effect of a chatbot tutor is stronger for women than for men. Therefore, chatbot may be a valid tool to lessen the gender gap in the information systems discipline.

1. Introduction

A recent report published jointly by Accenture and Girls Who Code indicates that the gender gap in the information technology industry may be worsening [1]. For example, women comprised 35 percent of IT workers in 1984 compared to 32 percent in 2020. In addition, women leave IT roles at a rate 45 percent higher than men. Part of this problem can be traced back to university programs where only 25 percent of tech graduates are women, who tend to drop IT classes at a higher rate than others (37 percent compared to 30 percent). This gap is problematic for both ethical/moral reasons and for organization performance given the ample evidence that greater diversity has a direct positive effect on performance outcomes in teams [2], top management [3], and organization level [4].

One explanation for this gap is that although there are not actual gender differences in the ability to learn IT skills and perform in IT roles, college and work cultures are not inclusive enough [1]. Research shows that while there are no performance differences between women and men, women tend to have lower confidence in their abilities—thus, leading them to avoid tech disciplines more so than men [5, 6]. As a result, research that can explain both why women feel less confident and strategies to reduce these feelings would be of great value to the information systems (IS) discipline.

A recent study by Garrow, et al. [5] demonstrated one possible explanation, or correlation, for gender confidence differences. In their study of dropout rates from an introductory IS programming course, they found that women are significantly less likely to seek help on their coursework from instructors, teaching assistants, fellow classmates, and even their own family members. Similar results have been found in organizational learning [7]. A common explanation for this difference is that help seeking implies incompetence or failure [7]. Given that women have been demonstrated to have higher average rate of self-consciousness [8] and greater fear of judgement and failure [9], it is understandable that they would be less likely to seek help from others who they feel may judge them. It is no surprise that women have been found to exhibit a higher rate of self-concealment during their university education [10].

Help seeking is a critical success factor in both university [11] and organizational learning [12]. Therefore, if the IS discipline is going to eliminate the gender gap in students, employees, and top management, it is imperative to help women feel comfortable seeking help without fear of judgement beginning in their early-stage coursework.

The root of a student's fear of social judgement comes from the negative emotions (shame, embarrassment, and lowered self-esteem) derived from the assumption that others (instructors, teaching

assistants, classmates, family, and friends) will harshly judge the student and rate her or him less competent or less important [9]. However, once a student feels competent, this fear subsides and is no longer a barrier to their continuance. Therefore, the purpose of this research is to identify and test a solution to reduce or eliminate this fear of judgement during the early stages of a college student's educational experience until they can gain sufficient self-confidence to ensure their continuance in the IS discipline.

Conceptually, there are two general approaches to address this issue. Either a) help women (and men) to overcome or eliminate the fear of judgement and self-consciousness, or b) provide help without a possibility for human judgement and shame. While both may be worthwhile strategies, this study focuses on the latter option by examining a fast-evolving technology from our discipline: chatbots—computer programs designed to simulate conversation with human users. Chatbots are increasingly used for both organizational and university learning [13]. Chatbots eliminate the concern of social judgement and shame by taking the human out of the equation. Chatbots have evolved significantly over time and offer an interesting degree of human-likeness which can make them a valid alternative to seeking help from another human [14].

For this study, we developed a custom chatbot for the Slack messaging platform to examine whether it can improve confidence among IS students in an introductory programming course at a large private university. We confirm two known findings: 1) women have less initial confidence than men, and 2) despite confidence differences, there is no performance difference between men and women. We also contribute two new findings: 3) a chatbot tutor can increase confidence among all students, and most importantly, can have a greater positive effect on women than men. The implication is that chatbots may be a promising strategy to retain women during their formative secondary coursework.

2. Emergence of Chatbots in Education

Chatbots have become a popular topic in education the last several years. Chatbots have been created to aid in learning foreign languages, answering administrative questions and frequently asked questions [13, 15-18]. There has also been an increase in the number of chatbots built to guide students in learning programming languages, especially in beginning level programming classes. For example, Hobert [19] built a chatbot that could respond to open-ended questions, assess source code, or guide students in step by step learning; Daud, et al. [20] created a chatbot to support students on Java programming problems; and Okonkwo and Ade-Ibijola [21] developed a chatbot for teaching Python to programmers to improve comprehension. Because of the early success of chatbots, researchers have studied ways of making them more effective such as demonstrated serverless chatbot architecture [22], including natural language processing and machine learning capabilities [18], improving chatbot efficiency [23], improving chatbot comprehension [21], and maximizing memory retention [24]. Overall, the vast majority of chatbots studied in prior research were found to be useful sources of help for students and improved student outcomes.

Chatbots can serve many purposes as an assistant in the classroom. In some ways, chatbots can perform better than a human teaching assistant (TA). A chatbot assistant in the classroom can be available 24/7 to answer student questions, always responds instantly, and will not judge or become frustrated with student queries. As mentioned above, students may be willing to ask questions of a chatbot that they would not ask a person for fear of embarrassment. Because of their conversational interface, a chatbot can act as a social companion, increasing engagement [14].

2.1. Why Women May Benefit More Than Men

Our primary motivation and intention is to discover whether women may benefit more than men when using chatbots in an education context.

While chatbots can be a resource available to all students, they are particularly beneficial for those who might be hesitant to approach someone else for help (such as a teaching assistant or instructor). That hesitancy may come from a variety of places, including a desire not to appear incompetent or unknowledgeable, or from general social anxiety [9]. For example, in their study of women in STEM (science, technology, engineering, and math) disciplines, Nelson, et al. [9] discovered that women had greater fear of shame, embarrassment, others devaluing their self-estimate, future uncertainty, and failure which led to lower levels of self-esteem and self-efficacy.

Providing an automated computer conversation with no human in the loop can help ease the burden for those who would not feel comfortable displaying vulnerability [25]. However, this logic only explains why a student may prefer a chatbot over a human for help. To understand gender differences, we turn to learning theory incorporating gender differences.

Severiens and Ten Dam [26] compared two of the predominant learning theories to derive an overall model that could specifically differentiate female and male characteristics. Their model is based on the *epistemological reflection theory* [27] on ways of knowing and patterns of reasoning as well as Vermunt's

theory [28] on learning styles and learning conceptions.

The epistemological reflection theory posits that men and women develop at similar rates, but with important differences in their process of progressing their knowledge from an absolute and factual information to relative and contextual theory-based wisdom. In a longitudinal study establishing this theory, Magolda [27] discovered that women are more open to learning from the perspectives and experiences of others which they incorporate into their own perspective whereas men are more individually focused on their own learning process and perspective. The implication of this is that women would be more naturally inclined to benefit from help seeking whereas men benefit more by learning on their own.

This epistemological reflection theory implies quite a tension for women who simultaneously feel more apprehensive about help seeking because of social judgment and embarrassment. Based on this theory, chatbots are a custom-fit solution for women because they simulate learning from the perspectives of another (as opposed to figuring things out on their own as favored by men) without the shame or embarrassment of feeling judged by another.

Furthermore, Magolda [27] identifies two general patterns of learning that include the roles of the learner, peers, teacher, assessments, and the nature of the knowledge. Adapted from [26, 27], Table 1 summarizes these roles and patterns:

Table 1. Epistemological reflection theory

	Receiving pattern	Mastering pattern
Role of learner: obtain knowledge	listening	ask questions
Role of peers: share knowledge	social contact, create relaxed atmosphere	debate learning material
Role of teacher: provide knowledge	minimal interaction	use interesting methods
Role of assessments: demonstrate knowledge	multiple opportunities	provide feedback
Nature of knowledge: certain or absolute	facts and opinions	value depends on degree of detail
Predominant gender of learners favoring the pattern	female	male

Magolda [27] found that women generally preferred the *receiving* pattern in which the learner is expected to listen as opposed to asking questions. Peers are expected to provide social contact and created a relaxed atmosphere. Teachers are expected to have minimal interaction. Assessments should provide multiple opportunities to succeed. By nature, those who

prefer the receiving pattern perceive knowledge as a set of facts and opinion.

This pattern, generally preferred by women (on average) more than the mastering pattern, is well-suited to chatbot technology. For example, in the classroom, students can learn by listening—knowing that they can ask questions later from the chatbot. Because the chatbot allows the learner to use peers for social contact rather than to debate learning material, they will have a more relaxed atmosphere in the classroom. The chatbot allows the learner to have minimal interaction with the instructor while interesting methods of learning can be programmed into the chatbot. A chatbot can have assessments integrated, but it may not have a direct effect on assessments at all. Finally, a chatbot is easy to program with facts and opinions although it can also include various levels of granularity in the details.

In summary, a chatbot allows the learner, peers, and teacher to fulfill their roles well regardless of whether the learner prefers the receiving pattern versus the mastering pattern; thus, making the educational experience more dynamic. Given that women generally prefer the receiving pattern, and assuming that IS courses are taught more commonly by men favoring the mastering pattern, we expect that chatbots will have a stronger effect on improving the learning experience of women than men.

2.2. How Women Will Benefit More Than Men

Several outcomes of help seeking have been identified and studied in prior literature including student engagement [29], knowledge increase [30], and self-efficacy [5]. We are particularly interested in self-efficacy. Based on Bandura's social learning theory [31], self-efficacy is one's confidence in their ability to perform an action or accomplish a task. Figure 1 visualizes the basic antecedents and consequences of self-efficacy [31].
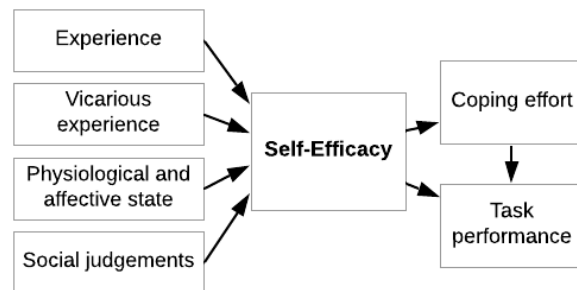

Figure 1. Self-efficacy theory

Self-efficacy theory posits that an individual's performance outcome for a given task may be based more on their *confidence* in their ability to perform the

task than their *actual* ability. This is because those with higher confidence in their capability are more likely to exhibit "coping" strategies and persist longer on tasks. As a result, an individual with high self-efficacy and low capability may outperform another with low self-efficacy but high capability.

Our self-efficacy varies over time based on our experiences, observations of others (vicarious experience), social judgements, and psychological states, our level of self-efficacy is theorized as a process model where it is both an indicator to, and outcome from, our performance over time. For example, self-efficacy has been modeled as both an indicator of help seeking in the classroom [32] (and, therefore, performance) as well as an outcome of help-seeking [5]. Self-efficacy has also be correlated with lower levels of fear, embarrassment, isolation, judgement, and discrimination in the classroom [32]. Self-efficacy also predicts student retention [5] which is the motivating factor of this study making it ideal for our purposes.

Self-efficacy has been modeled both as a global, reflective construct as well as task-specific formative construct [33]. We are interested in task-specific self-efficacy because we wish to model the effects of chatbots (moderated by gender) on a student's confidence that they can perform the tasks taught in a particular course. Task self-efficacy must be defined uniquely to the context it is relevant to. Based on the unique purposes of the IS discipline, we draw from research on creative self-efficacy (CreaSE).

The belief in one's ability to creatively solve unstructured problems is referred to as CreaSE [Gong et al. 2009; Richter et al. 2012; Tierney and Farmer 2002]. Most recently, IS researchers have adapted this construct to measure IS creative self-efficacy (CreaSE) which refers to, "…an individual's belief in their ability to develop creative solutions to new or unstructured business problems through the development of information systems that support business process and the people who execute them" [34, pg. 5].

Based on core theory on human creativity [35], CreaSE is a second-order formative construct with five independent factors [34]: 1) affect, 2) business skills, 3) intelligence, 4) people skills, and 5) technology training. *Affect* refers to our emotions, moods, and attitude [36] toward creative problem solving which has a significant effect on our creative performance. For example, negative affect can reduce our "flexible thinking" and problem-solving capabilities on complex tasks [37]. *Business skills* is a person's knowledge about the business domain they are working in including processes, strategies, and management. *Intelligence* refers to the cognitive ability for creativity a person is innately born with. *People skills* is a person's ability to collaborate effectively with others on a team and

combine the good ideas from others into their own problem-solving framework. Finally, *technology training* refers to the technology skills, such as programming, data analytics, and computer systems, which will be combined and implemented in creative ways to solve IS problems (see Figure 2).
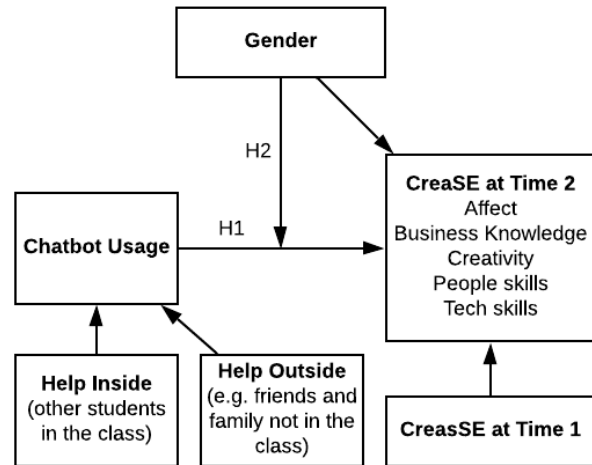

Figure 2. Self-efficacy theory

In summary, we model self-efficacy using the CreaSE measurement and model chatbot usage as an antecedent of CreaSE. Chatbot usage is a form of *mastery of experience* as defined in self-efficacy theory [31]. In particular, successful usage of the chatbot to find answers and complete course assessments should lead to greater confidence. To understand the effects of chatbot usage on CreaSE, it is important to measure CreaSE over time. Therefore, we include a CreaSE measure before chatbot usage as an important covariate.

To validate whether chatbot usage is truly replacing the need for help in coursework, we include two antecedents of chatbot usage: help received from peers, instructor, and TAs for the course, and help received from outside of the course such as family members and friends who are not currently enrolled in the course.

Because women have been found to have significantly lower self-efficacy than men in both academic [38] and technology [39] contexts, we model gender as a covariate explaining CreaSE. The two relationships that are unique to our study are hypothesized below:

> **H1:** *Use of a chatbot teaching assistant increases the CreaSE of all students*

> **H2:** *Use of a chatbot teaching assistant increases the CreaSE of women more than men*

3. Methodology

To test our model, we developed a chatbot for the Slack platform (the primary communication medium used in the course) for students at a private university in the western United States. Students participated in a pre- and post-test survey at the beginning of the Spring 2021 semester. Both the surveys included measures of CreaSE [5] using a 7-point Likert scale as well as a test on coding knowledge and skills based on prior research designed to accurately measure student programming knowledge [40]. Although not included in our model, we measured a pre- and post-test of programming knowledge to confirm that women and men have equal performance regardless of self-efficacy differences.

Demographics were measured in the post-test (see Appendices A and B). A total of 147 students enrolled in the intro to computer programming course across four different sections with three different instructors. Of those, 136 completed both the pre- and post-test survey in full. The average age of survey participants was 21.5 years old with just under 73% being male and the other 27% being female. After the pre-test survey was completed, students were introduced to the chatbot and made aware that they could use it, if they desired, to help them in the course. The students were given approximately seven weeks to interact with the chatbot following which they completed the same survey again. This post-test survey also included a self-reported measure of the student's perception of how much they used the chatbot during those seven weeks. Survey data was dropped for students that failed to complete the pre-chatbot usage or post-chatbot usage surveys as the data could not be used to draw any conclusions.

3.1. Chatbot Details

The chatbot, which we named Skylar, was built using Google's DialogFlow as the conversational platform, MySQL for the database, and Slack as the connection platform where students could directly interact with the chatbot. The name Skylar was chosen to be gender-neutral. Skylar has the following four main functions: help students learn coding concepts (learn function), drill students' coding skills by asking coding questions based on a chosen topic (drill function), review for tests (review function), and provide practice programs for students to build in their preferred coding environment (practice program function). In addition to these functions, Skylar has some small talk capabilities including telling jokes. Skylar was already in the class Slack workspace so students only had to direct message the bot to get started; starting with "Hello," Skylar then guided them through how to use it by explaining the purpose and offering a main menu where the user could select one of the four main functions.
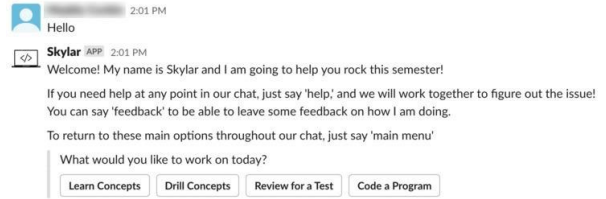

Figure 3. Starting conversations with skylar

3.1.2. Database. The MySQL database, used as a foundation for Skylar's response, includes descriptions, syntax, examples, and drill questions for each programming concept taught in the course (i.e., arrays, classes, variables, etc.). For each question, the database also contains the appropriate answer. Additional resources such as YouTube videos and links to W3Schools lessons students can reference in their learning are also included.

A user table keeps track of the unique ID of each Slack user which we used to trace each student's CreaSE and skills performance with his/her chatbot interaction.

3.1.3. Learn Function. When students choose the Learn Concepts button, they are asked to select a specific concept. The concept is taken from the request, then queried from the database and the explanation is returned to the user. Most concepts contain a definition of the concept, what the correct syntax looks like, and an example. In addition to this, students could select a "More Resources" button where additional resources were shared such as W3Schools links, class videos, or helpful resources from other sites. Students could also click the "Test My Knowledge" button to get 2 to 5 practice questions based on the specified concept.
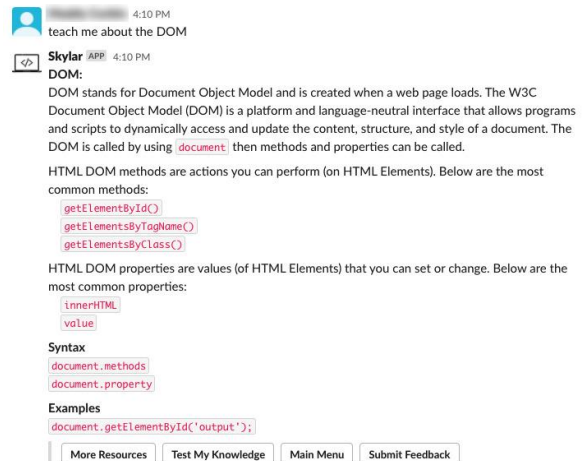

Figure 4. Learn function example

3.1.4. Drill Function. When students choose to drill their skills, 17 buttons appear covering all the main topics taught in the course. Students could pick one of

the options, then specify how many questions they wanted. Skylar then pulled related questions in random order from the database. Students could answer the question, then for the majority they could also view an explanation of concepts and mechanics to aid understanding. There are 3 types of questions including multiple choice, true or false, and open response questions. By having the different types, it allows for students to increase familiarity with conceptual aspects as well as practice writing code. After the quiz, Skylar gave students their score and study suggestions.
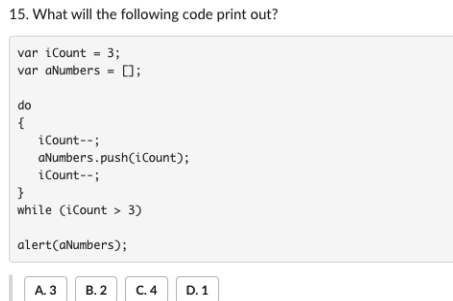
```
15. What will the following code print out?

var iCount = 3;
var aNumbers = [];

do
{
    iCount--;
    aNumbers.push(iCount);
    iCount--;
}
while (iCount > 3)

alert(aNumbers);
```

A. 3   B. 2   C. 4   D. 1

Figure 5. Drill function example

Each programming principle is labeled as a subcategory under a main concept in the database. For example, string methods (i.e., splice, indexOf, substring, etc.) all fall under the string concept. Each question is linked to a subcategory, and consequently, a main category as well. If a student is drilling the string concept, they may understand the splice method well but miss the questions on the substring method, in which case Skylar would inform the user to study substrings.

3.1.5. Review Function. The course has two midterms and a final exam. Every concept in the database was connected to an exam where students would be tested on the concept. When students chose the review function, they were asked which test they wished to study for and how many questions they would like. This information was then used to query sample questions from the database matching the practice exam chosen (for the final exam, it gathered questions from all categories).
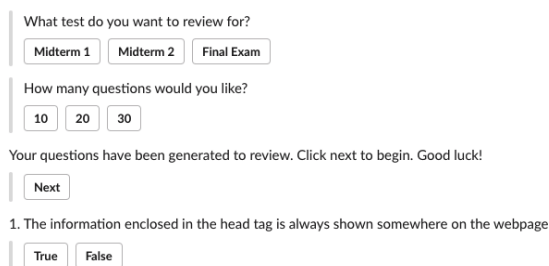
What test do you want to review for?
Midterm 1   Midterm 2   Final Exam

How many questions would you like?
10   20   30

Your questions have been generated to review. Click next to begin. Good luck!
Next

1. The information enclosed in the head tag is always shown somewhere on the webpage
True   False

Figure 6. Review function example

The drill and review function questions were pulled from the same table. At the end of the review students are given their score and concepts to study.

3.1.6. Practice Program Function. The practice program function was designed for students who want direction for extra mini-projects they could use to further test and develop their skills.
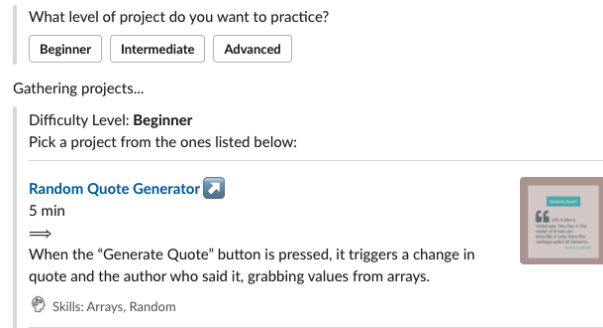
What level of project do you want to practice?
Beginner   Intermediate   Advanced

Gathering projects...

Difficulty Level: **Beginner**
Pick a project from the ones listed below:

Random Quote Generator
5 min
⟹
When the "Generate Quote" button is pressed, it triggers a change in quote and the author who said it, grabbing values from arrays.

Skills: Arrays, Random

Figure 7. Review function example

Students could choose from 18 programming prompts (six each for beginner, intermediate, and advanced levels). When this function starts, Skylar asks what level of program they want. After selecting the level, students see a list of prompts, including a title that links to more details if available, time estimate, picture, description, as well as a list of skills the program will use. These projects are coded in the users preferred coding environment.

3.5. Small Talk Ability. Skylar also has some small talk abilities that improve the experience which have been shown to be useful in chatbot research [19].

4. Results

4.1. Measurement Model

Table 2 summarizes the measures of convergent and discriminant validity, reliability of the latent constructs as well as the multicollinearity of all exogenous variables. All established criterion were met or exceeded [41]. The average variance extracted (AVE) for each latent factor were each above 0.50, suggesting that more than 50% of variation in levels of CreaSE in the surveyed IS students can be explained by chatbot usage. Composite reliability (C.R.) and Cronbach's alpha were above the 0.7 threshold. Every AVE—found in bold font along the matrix diagonal--was greater than the squared correlation. Although not depicted in Table 2, every variance inflation factor (VIF) for all variables was lower than 3.

Table 2. Measurement model statistics

	AFF	BUS	INT	PEO	TRA	α	C.R.
AFF	**0.817**	0.240	0.352	0.161	0.182	0.757	0.855
BUS	0.490	**0.799**	0.365	0.347	0.667	0.811	0.876
INT	0.593	0.604	**0.715**	0.292	0.292	0.759	0.836
PEO	0.401	0.589	0.540	**0.765**	0.407	0.763	0.849
TRA	0.427	0.817	0.540	0.638	**0.738**	0.723	0.825

4.2. Hypothesis Testing

The theoretical model (see Figure 2) was tested using a partial least squares (PLS) structural equation model (SEM) using the tool SmartPLS 3.2.8 [Ringle et al. 2015]. Path coefficients indicate the measured relationships. A bootstrapping procedure (3000 sub-samples) was used to estimate path significance.

Before testing our model, we generated another model to determine whether gender or chatbot usage had any effect on the student's actual performance as measured by the pre- and post-test programming skills questions found in the appendix. CODE refers to the five multiple choice questions while PSEUDO refers to the five questions requiring the students to write pseudo code to solve a problem. PSUEDO was scored/graded by two teaching assistants who agreed on all scores. We found no significant effect of gender on their performace on the programming knowledge questions answered at Time 2 (the end of the course). In addition, the use of the Skylar chatbot did not signficantly affect student performance on the programming questions. However, as expected, their performance on the pre-test programming questions did significantly impact their performance at the end of the semester (see Table 3).

Table 3. Coefficient estimates on performance

Relationship	β	t	P
Gender -> CODE	-0.110	0.919	0.179
Gender -> PSEUDO	-0.056	0.577	0.282
Skylar Minutes -> CODE	-0.161	1.008	0.157
Skylar Minutes -> PSEUDO	0.044	0.545	0.293
CODE: Time1 -> Time2	0.288	2.547	**0.006**
PSEUDO: Time1 -> Time2	0.644	10.196	**0.000**

After confirming that there are no performance differences by gender, nor by using the chatbot, we proceeded to test our theoretical model to see whether chatbot usage and gender had effects on CreaSE. Figure 8 vizualizes the results while Table 4 summarizes the exact coefficients, t-scores, and p-values. In summary, every subdimension of CreaSE improved significantly from Time1 (pre-chatbot) to Time2 (post-chatbot). The number of minutes estimated by the student that used the Skylar chatbot significantly improved the CreaSE sub-dimensions of affect (AFF), intellect (INT), and technology training (TRA). Business skills (BUS) and

people skills (PEO) were not significantly affected. Therefore, H1 was confirmed as three of the five dimensions were improved. In addition, help seeking from peers, TAs, and the instructor significantly reduced number of minutes spent using the chatbot (β = -0.154*). As expected, men (coded 0) had more CreaSE at Time2 than women (coded 1) which is consistent with prior research. Most importantly, the interaction terms for the CreaSE subdimensions AFF and INT were significantly higher for women than for men. Thus, H2 was confirmed, indicating that chatbots have a greater impact on improving confidence in women than men.
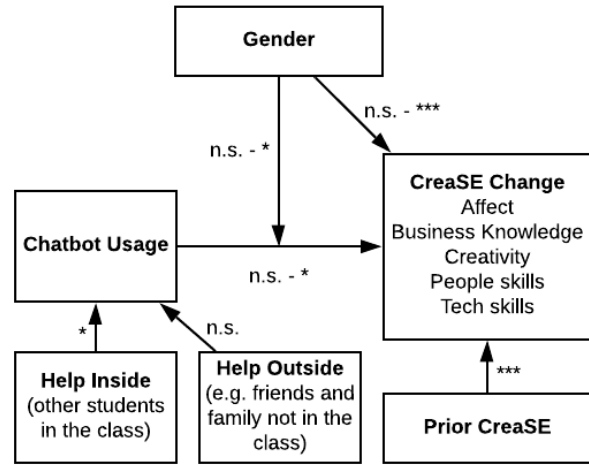

Figure 8. Model testing results

Table 4. Hypothesis testing results

Relationship	β	t	p
CreaSE_AFF: Time1 -> Time2	0.509	5.848	**0.000**
CreaSE_BUS: Time1 -> Time2	0.364	4.305	**0.000**
CreaSE_INT: Time1 -> Time2	0.426	4.772	**0.000**
CreaSE_PEO: Time1 -> Time2	0.560	7.021	**0.000**
CreaSE_TRA: Time1 -> Time2	0.484	6.205	**0.000**
Help_Inside -> Skylar Minutes	-0.154	1.897	**0.029**
Help_Outside -> Skylar Minutes	-0.019	0.211	0.416
Skylar Minutes -> CreaSE_AFF	0.172	2.092	**0.018**
Skylar Minutes -> CreaSE_BUS	0.040	0.356	0.361
Skylar Minutes -> CreaSE_INT	0.167	2.089	**0.018**
Skylar Minutes -> CreaSE_PEO	0.103	1.147	0.126
Skylar Minutes -> CreaSE_TRA	0.179	2.004	**0.023**
Gender -> CreaSE_AFF	-0.043	0.464	0.321
Gender -> CreaSE_BUS	-0.111	1.147	0.126
Gender -> CreaSE_INT	-0.273	3.112	**0.001**
Gender -> CreaSE_PEO	-0.186	2.311	**0.011**
Gender -> CreaSE_TRA	-0.140	1.738	**0.041**
Gender * Skylar min -> CreaSE_AFF	0.181	1.961	**0.025**
Gender * Skylar min -> CreaSE_BUS	0.076	0.633	0.264
Gender * Skylar min -> CreaSE_PEO	0.105	0.983	0.163
Gender * Skylar min -> CreaSE_TRA	0.082	0.826	0.205
Gender * Skylar min -> CreaSE_INT	0.170	1.624	**0.052**

Notes: 0=Male, 1=Female

5. Discussion

Each of the primary objectives of this research were accomplished and/or supported. First, our study confirmed two results from prior research: 1) women have less confidence than men in some IS course topics, and 2) despite confidence differences, there is no performance difference between men and women. We also contributed two unique findings: 3) chatbots can increase confidence (measured as CreaSE in the IS discipline) among all students (H1), and most importantly, 4) chatbots have a greater positive effect among women than men (H2).

Assuming that theories like *epistemological reflection* [27] accurately reflect the differences in learning process and orientation between women and men, then chatbots appear to be a valid and useful tool to help women in IS courses to develop greater confidence which can be replicated beyond this research. By so doing, chatbots may help IS programs to retain more women who begin foundational courses in our discipline as noted by Garrow, et al. [5]; thus, helping to reduce the gender gap in female representation in IS programs.

There are several specific findings of interest that should be discussed in greater detail. For example, men's levels of confidence were higher than that of women which is what we expected based on prior research [38]. This validates the course context used for this study since it replicates that of known research. However, it is interesting to note that this was only for intellect, people skills, and technology training—not for attitude or business skills. This highlights the unique problem that STEM fields like IS face. Other disciplines like general business do not face the same gender confidence gap that IS does.

Three of the five subdimensions of CreaSE (affect, intellect, and technical training) increased over the course of the semester with chatbot interaction. However, business skills and people skills were not affected. This is explainable because the purpose of the course (and chatbot) was to improve programming skills (i.e. technical training) which included the need to be creative to complete course assignments. Successfully completing these tasks would naturally increase one's confidence in their attitude, intellect, and training. It is logical that business skills and people skills would be unaffected. Therefore, the lack of effect on those sub-dimensions does not detract from supporting H1.

The significant effect of getting help from others within the course (peers, TAs, and instructor) on minutes spent using the chatbot confirms that the chatbot was a valid option for help seeking. The fact that help sought outside of the course (family and friends not enrolled) did not significantly reduce the minutes spent using the chatbot does not invalidate the chatbot, but simply indicates that outside help was not needed.

Also of note, the greater improvement in CreaSE that women found in the affect and technical training sub-dimensions (thus, supporting H2) were not replicated in the other three sub-dimensions. However, the coefficients for those three moderating effects were positive. Therefore, if those relationships are replicated in future research, they would eventually be statistically significant with a large enough sample size. The smallest of those coefficients was the moderation of gender on the effect of chatbot usage on business skills ($\beta = 0.076$)—again highlighting that the need to help women gain confidence is not as great as the general business discipline.

The primary limitation of our study was the measure of chatbot usage. It was a self-reported perception measure of how many minutes they spent using the chatbot. Although the student could view their chatbot history in making this estimation, it will naturally have some error. We propose that future research conducts similar experiments with a more valid measure of time spent with the chatbot (e.g. usage logs).

Future research should also verify whether the improved confidence in women from chatbots is maintained throughout future courses and whether it truly improves retainment of women in IS programs.

6. References

[1] Accenture, and Girlswhocode, https://www.accenture.com/_acnmedia/PDF-136/Accenture-Resetting-Tech-Culture.pdf, June 6th, 2021, 2020

[2] Anderson, G., Keith, M., Albrecht, C., Spruill, A., and Pettit, C., "Optimizing Software Team Performance with Cultural Differences", Hawaii International Conference on System Sciences, 2019

[3] Ruiz-Jiménez, J.M., Del Mar Fuentes-Fuentes, M., and Ruiz-Arroyo, M., "Knowledge Combination Capability and Innovation: The Effects of Gender Diversity on Top Management Teams in Technology-Based Firms", Journal of Business Ethics, 135(3), 2016, pp. 503-515.

[4] Zhang, L., "An Institutional Approach to Gender Diversity and Firm Performance", Organization Science, 31(2), 2020, pp. 439-457.

[5] Garrow, P., Anderson, G., Keith, M.J., and Spruill, A.N., "Student Retention in Information Systems Majors: The Role of Creative Self-Efficacy": Book Student Retention in Information Systems Majors: The Role of Creative Self-Efficacy, 2019, pp. 10.

[6] Duran, R., Haaranen, L., and Hellas, A., "Gender Differences in Introductory Programming: Comparing Moocs and Local Courses": Book Gender Differences in Introductory Programming: Comparing Moocs and Local Courses, 2020, pp. 692-698.

[7] Lee, F., "When the Going Gets Tough, Do the Tough Ask for Help? Help Seeking and Power Motivation in

Organizations", Organizational Behavior and Human Decision Processes, 72(3), 1997, pp. 336-363.

[8] Rankin, J.L., Lane, D.J., Gibbons, F.X., and Gerrard, M., "Adolescent Self-Consciousness: Longitudinal Age Changes and Gender Differences in Two Cohorts", Journal of Research on Adolescence, 14(1), 2004, pp. 1-21.

[9] Nelson, K.L., Newman, D.N., Mcdaniel, J.R., and Buboltz, W.C., "Gender Differences in Fear of Failure Amongst Engineering Students", International Journal of Humanities and Social Science, 3(16), 2013, pp. 10-16.

[10] Morgan, N.T., and Robinson, M., "Students' Help-Seeking Behaviours by Gender, Racial Background, and Student Status", Canadian Journal of Counselling, 37(2), 2003, pp. 151-166.

[11] Karabenick, S.A., "Seeking Help in Large College Classes: A Person-Centered Approach", Contemporary educational psychology, 28(1), 2003, pp. 37-58.

[12] Grodal, S., Nelson, A.J., and Siino, R.M., "Help-Seeking and Help-Giving as an Organizational Routine: Continual Engagement in Innovative Work", Academy of Management Journal, 58(1), 2015, pp. 136-168.

[13] Winkler, R., and Soellner, M., "Unleashing the Potential of Chatbots in Education: A State-of-the-Art Analysis", 2018,

[14] Schuetzler, R.M., Grimes, G.M., and Scott Giboney, J., "The Impact of Chatbot Conversational Skill on Engagement and Perceived Humanness", Journal of Management Information Systems, 37(3), 2020, pp. 875-900.

[15] Kerlyl, A., Hall, P., and Bull, S., "Bringing Chatbots into Education: Towards Natural Language Negotiation of Open Learner Models": Book Bringing Chatbots into Education: Towards Natural Language Negotiation of Open Learner Models, Springer, 2006, pp. 179-192.

[16] Molnár, G., and Szüts, Z., "The Role of Chatbots in Formal Education": Book The Role of Chatbots in Formal Education, IEEE, 2018, pp. 000197-000202.

[17] Chocarro, R., Cortiñas, M., and Marcos-Matás, G., "Teachers' Attitudes Towards Chatbots in Education: A Technology Acceptance Model Approach Considering the Effect of Social Language, Bot Proactiveness, and Users' Characteristics", Educational Studies, 2021, pp. 1-19.

[18] Pérez, J.Q., Daradoumis, T., and Puig, J.M.M., "Rediscovering the Use of Chatbots in Education: A Systematic Literature Review", Computer Applications in Engineering Education, 28(6), 2020, pp. 1549-1565.

[19] Hobert, S., "Say Hello to 'Coding Tutor'! Design and Evaluation of a Chatbot-Based Learning System Supporting Students to Learn to Program", 2019,

[20] Daud, S.H.M., Teo, N.H.I., and Zain, N.H.M., "E-Java Chatbot for Learning Programming Language: A Post-Pandemic Alternative Virtual Tutor", International Journal, 8(7), 2020,

[21] Okonkwo, C.W., and Ade-Ibijola, A., "Python-Bot: A Chatbot for Teaching Python Programming", Engineering Letters, 29(1), 2020,

[22] Yan, M., Castro, P., Cheng, P., and Ishakian, V., "Building a Chatbot with Serverless Computing": Book Building a Chatbot with Serverless Computing, 2016, pp. 1-4.

[23] Santana, A.D.O., and Aranha, E., "An Approach to Generate Virtual Tutors for Game Programming Classes": Book An Approach to Generate Virtual Tutors for Game Programming Classes, 2019, pp. 246-252.

[24] Abbasi, S., and Kazi, H., "Measuring Effectiveness of Learning Chatbot Systems on Student's Learning Outcome and Memory Retention", Asian Journal of Applied Science and Engineering, 3(2), 2014, pp. 251-260.

[25] Kang, S.H., and Gratch, J., "Virtual Humans Elicit Socially Anxious Interactants' Verbal Self-Disclosure", Computer Animation and Virtual Worlds, 21(3-4), 2010, pp. 473-482.

[26] Severiens, S., and Ten Dam, G., "Gender and Learning: Comparing Two Theories", Higher Education, 35(3), 1998, pp. 329-350.

[27] Magolda, M.B.B., Knowing and Reasoning in College: Gender-Related Patterns in Students' Intellectual Development, Jossey-Bass, 1992.

[28] Vermut, J.D., "Metacognitive, Cognitive and Affective Aspects of Learning Styles and Strategies: A Phenomenographic Analysis", Higher Education, 31(1), 1996, pp. 25-50.

[29] Caraway, K., Tucker, C.M., Reinke, W.M., and Hall, C., "Self-Efficacy, Goal Orientation, and Fear of Failure as Predictors of School Engagement in High School Students", Psychology in the Schools, 40(4), 2003, pp. 417-427.

[30] Mäkitalo-Siegl, K., Kohnle, C., and Fischer, F., "Computer-Supported Collaborative Inquiry Learning and Classroom Scripts: Effects on Help-Seeking Processes and Learning Outcomes", Learning and Instruction, 21(2), 2011, pp. 257-266.

[31] Bandura, A., "Self-Efficacy: Toward a Unifying Theory of Behavioral Change", Psychological Review, 84(2), 1977, pp. 191-215.

[32] Maeda, J., "Self-Efficacy Reduces Impediments to Classroom Discussion for International Students: Fear, Embarrassment, Social Isolation, Judgment, and Discrimination", IAFOR Journal of Education, 5(2), 2017, pp. 141-159.

[33] Marakas, G.M., Yi, M.Y., and Johnson, R.D., "The Multilevel and Multifaceted Character of Computer Self-Efficacy: Toward Clarification of the Construct and an Integrative Framework for Research", Information systems research, 9(2), 1998, pp. 126-163.

[34] Payne, K.C., Keith, M.J., Babb, J., and Spruill, A.N., "Development and Validation of the Information Systems Creative Self-Efficacy Scale": Book Development and Validation of the Information Systems Creative Self-Efficacy Scale, Hilton Waikoloa Village, Hawaii, 2018

[35] Hennessey, B.A., and Amabile, T.M., "Creativity", Annual Review of Psychology, 61(2010, pp. 569-598.

[36] Blanchette, I., and Richards, A., "The Influence of Affect on Higher Level Cognition: A Review of Research on Interpretation, Judgement, Decision Making and Reasoning", Cognition & Emotion, 24(4), 2010, pp. 561-595.

[37] Aspinwall, L.G., "Rethinking the Role of Positive Affect in Self-Regulation", Motivation and emotion, 22(1), 1998, pp. 1-32.

[38] Huang, C., "Gender Differences in Academic Self-Efficacy: A Meta-Analysis", European journal of psychology of education, 28(1), 2013, pp. 1-35.

[39] He, J., and Freeman, L.A., "Are Men More Technology-Oriented Than Women? The Role of Gender on the Development of General Computer Self-Efficacy of College

Students", Journal of Information systems education, 21(2), 2010, pp. 203-212.

[40] Oliver, R., "Measuring Hierarchical Levels of Programming Knowledge", Journal of educational computing research, 9(3), 1993, pp. 299-312.

[41] Fornell, C., and Larcker, D.F., "Evaluating Structural Equation Models with Unobservable Variables and Measurement Error", Journal of Marketing Research, 18(1), 1981, pp. 39-50.

7. Appendix

Table A1. CreaSE survey questions

AFF1: I believe that I would have the determination necessary to creatively solve IS problems.
AFF2: I believe that creative IS problem solving would be very interesting to me.
AFF3: I believe that I would be enthusiastic about creative IS problem solving.
BUS1: I believe that I have the domain knowledge necessary to identify the root cause of an IS problem.
BUS2: I believe that I have the expertise necessary to solve complex IS problems.
BUS3: I believe that I have enough knowledge about business processes to create better solutions.
INT1: I believe that I have the ability to understand an IS problem from multiple angles.
INT2: I believe that I have the ability to think unconventionally to find IS solutions.
INT3: I believe that I have the mental capacity required to understand the root cause of an IS problem.
PEO1: I believe that I understand why people respond the way they do to new IS ideas.
PEO2: I believe that I understand how people will react to changes in business processes.
PEO3: I believe that I understand how people will interact with new technology.
TRA1: I believe that I have the data collection skills necessary to create IS solutions.
TRA2: I believe that I have the data analysis training required to create unique IS solutions.
TRA3: I believe that I have the IT skills necessary to create new IS solutions.

Table A2. Competence survey questions

C1	Which of the following are characteristics of a strongly typed programming language? Mark all that apply. A. In strongly typed languages, data types must be declared for each variable defined. B. Variables are declared without defining which data type the variable is in strongly typed languages. C. When using a strongly typed language, only the keywords const and let can be used to create variables. D. Errors are less likely to occur during compile time when using a strongly typed language as opposed to a loosely typed language.
C2	What is the difference between a function and a method in programming?
C3	You're given a list of textbooks with their respective prices and need to add up the total cost of all the textbooks. Which of the following would be the BEST statement to use? A. If/then statement
	B. Switch statement C. For loop D. While loop
C4	A class in programming is a prototype that can be used to create objects. Describe below what an object is:
C5	Which of the following correctly describes characteristics of a compiled programming language? Mark all that are true. A. In compiled languages, a second program is required to understand and execute the code. B. A processor can directly execute the code from a compiled language. C. Compiled languages are generally faster than interpreted languages. D. Compiled languages must be rebuilt after every change.
PC 1	Write the logic in pseudocode to see if baked cookies are done. Take into consideration a time and the level of doneness for the cookies. If the cookies are done, take them out of the oven.
PC 2	Write the logic in pseudocode to increment the score of a basketball game. Only worry about the home team score. If the team makes a free throw, add 1 point to the score. If they make a regular shot, add 2 points; and if they make a 3-pointer, add 3 points.
PC 3	Write the logic in pseudocode to count how many items you have in your grocery basket to get a total price.
PC 4	Write the logic in pseudocode for a block of code you want to use in multiple programs. The code should get student information including first name, last name, student ID number, major, and GPA.
PC 5	You have a list of teams with a team name and team score for each team. Write the pseudocode that sorts the list of teams with their scores ranked from lowest to highest and maintains the team name with the appropriate score.