Understanding Students’ Acceptance of Online Judge System in Programming Courses: A Structural Equation Modelling Approach

DARAMSENGE BILEGJARGAL, NIEN-LIN HSUEH
Department of Information Engineering and Computer Science, Feng Chia University, Taichung 40724, Taiwan

Corresponding author: Daramsense Bilegjargal (p0731809@o365.fcu.edu.tw).

This research was supported by the Ministry of Science and Technology, Taiwan, under grant MOST 109-2221-E-035-056.

ABSTRACT An online judge (OJ) system was developed to evaluate programs in online programming contests. They have also been widely applied to help students practice their coding skills; however, no studies have investigated the acceptance of OJ technology by students in online programming courses. In this study, we applied the second generation of the unified theory of acceptance and use of technology (UTAUT2) model and partial least-squares structural equation modeling (PLS-SEM) to fill this research gap. We recruited 187 undergraduate participants from the Data Science course at Feng Chia University (FCU), Taiwan, in the spring semester of 2021. Our results showed that ‘hedonic motivation’, ‘self-efficacy’, and ‘social influence’ had the most significant positive effects on students’ intention to use the OJ system, and ‘anxiety’ negatively impacted the intention to use the OJ system. Our results can serve as a reference for OJ system developers, designers, instructors, and policymakers within universities.

INDEX TERMS Online judge system, programming education, PLS-SEM, UTAUT2, technology acceptance, learning behavior.

I. INTRODUCTION
Programming assignments are important practical tasks in engineering degree programs. However, manual assessment of these assignments is tedious and time-consuming for teachers [1]. Therefore, automated programming assessment systems (APAS) have been developed to track the learning progress of students and reduce the workload of educators [2]. For example, Mekterovic et al. [3] developed ‘Edgar’ - a novel, state-of-the-art, automated program assessment system capable of assessing (1) programming assignments written in arbitrary programming languages (e.g., SQL, Java, C, C#) as well as (2) multi-correct multiple-choice questions. ‘Edgar’ features a flexible exam and question definition model, rich exam logging and monitoring facilities, as well as data analysis and visualization features.

An online judge (OJ) is a web-based software that was developed to judge uploaded source code in online programming contests [4]. Judging involves compiling source codes into executables, running them against sets of input data, and comparing the results with standard results [5]. Widely known OJs include Codeforces, URI Online Judge, Uva, and SPOJ [6]. However, challenges exist when applying OJ systems to online programming courses. For example, OJs place the responsibility for learning on the student, who may lack motivation or appropriate study habits [7]. While many studies have investigated how to improve the design of OJ systems for programming courses, research into student acceptance of OJ systems is lacking. We aimed to fill this gap by exploring the following research questions:

- What are the significant factors that affect students’ acceptance of and intention to use an OJ system in an online programming course?
- Does academic major moderate the relationships between selected candidate factors and behavioral intention to use an OJ system?

To answer these research questions, we applied the second generation of the unified theory of acceptance and use of technology (UTAUT2) model and partial least-squares structural equation modeling (PLS-SEM). We recruited participants from the ‘Foundations and Application of Data Science’ course at Feng Chia University, Taiwan, during the six weeks of the spring semester of 2021. Our results provide valuable references.
for the application and development of OJ systems, as well as for online programming course design.

The remainder of this paper is organized as follows. Section 2 discusses previous studies on OJ systems, UTAUT, and PLS-SEM. Sections 3 and 4 present the theoretical background and research methodology, respectively. Sections 5 and 6 present the results of our data analysis, relevant discussions, and implications. Finally, Sections 7 and 8 describe the research limitations, future work, and conclusions, respectively.

II. LITERATURE REVIEW

A. ONLINE JUDGE (OJ) SYSTEMS

OJs are tools for the automatic grading of programming assignments that have been widely used to support computer science (CS) courses. While much research has investigated means of improving OJ systems for programming tasks, few studies have used these tools to acquire and analyze interaction data to better understand student behaviors. This gap can be attributed to a lack of data availability or inadequate granularity [8].

Petit et al. [9] previously reported on the use of Jutge.org, a free OJ system, in a wide range of courses that covered basic programming, data structures, algorithms, artificial intelligence, functional programming, and circuit design at the Universitat Politecnica de Catalunya. For ten years, the Jutge.org tool has played an essential role in improving programming courses. The survey results published by Petit et al. [9] showed that most students consider the OJ to be of great help. Furthermore, tournament activities based on this system were described as ‘fun, attractive, and motivating’.

Aleman [10] proposed extending the use of OJs with a series of assignments related to programming tools. He conducted an experiment that compared the effects of automated assessments with those of conventional face-to-face classroom teaching. In particular, Aleman [10] was interested in the effects on the acquisition and retention of versioning skills and knowledge, as well as on testing, debugging, and deployment. His results indicated that automated assessment systems can promote student interest and produce statistically significant differences in the scores of the experimental and control groups.

Galan et al. [11] developed a framework to automatically review programming assignments and reported eight years of data collected using this framework. Their results indicated that OJ systems were accepted by students after proper training, and that student grades were strongly correlated with assignment completion.

B. UTAUT

Venkatesh et al. [12] proposed the UTAUT, a model that consists of four core variables (performance expectancy, effort expectancy, social influence, and facilitating conditions) and four moderating variables (gender, age, experience, and voluntariness of use). The UTAUT model integrates eight major theories of technology acceptance, and its efficacy has been verified using a large, real-world dataset. Later, the UTAUT model was extended to UTAUT2 by incorporating three new constructs: hedonic motivation, price value, and habit [13], as shown in Fig. 1.

Huang [17] adapted UTAUT2 to investigate the factors that influence user acceptance of fully immersive VR compared to desktop VR. His results, obtained from 56 participants, indicated that (1) performance expectancy, hedonic motivation, and facilitating conditions were useful predictors of behavioral intentions to use VR systems, and (2) both types of VR systems were well accepted by users.

Jakkaw and Hemrungrote [18] employed the UTAUT2 model to explore factors that determine the deployment of Google Classroom in a course entitled ‘Introduction to Information Technology’ at Mae Fah Luang University in Chiang Rai, Thailand. Their findings confirmed that performance expectancy, effort expectancy, social influence, and facilitating conditions all influence behavioral intentions to use Google Classroom. However, although students agreed that Google Classroom is a useful and user-friendly tool, most of its features are not applied at their full capability.

Thong et al. [19] extended UTAUT to the context of information and communication technology (ICT) services by examining the moderating role of ICT service type. In so doing, they tested the proposed model in a large-scale survey of 4777 consumers and found strong empirical support for their model. Specifically, those researchers found that (1) service type moderated key relationships and (2) the moderated model explained a large proportion of the variance (50%–66%) in behavioral intentions to use ICT services.

Salloum et al. [20] examined student acceptance of e-learning at five universities in the United Arab Emirates. Results obtained from a total of 435 students indicated that system quality, computer self-efficacy, and computer playfulness significantly impacted the perceived ease of use of e-learning systems. Information quality, perceived enjoyment, and accessibility were also found to positively influence the perceived ease of use and usefulness of e-learning systems.

The UTAUT2 model consolidates factors that determine how people use technology, including digital platforms [21]. The UTAUT2 model incorporates the theory of reasoned action (TRA), the theory of planned behavior (TPB), and the technology acceptance model (TAM). Therefore, it is the most comprehensive choice for our research aims and should have strong explanatory power [22].

C. PLS-SEM

Over the years, educational researchers have used statistical methods to extend their abilities to develop, explore, and confirm research findings. PLS-SEM is a widely used method [14]. This multivariate statistical modeling technique is a nonparametric bootstrap method that makes
no distributional assumptions and can be estimated with small sample sizes [15]. It represents a causal modeling approach that aims to maximize the explained variance of dependent latent constructs. PLS-SEM path modeling, if appropriately applied, is considered to be a “silver bullet” for estimating causal models in many contexts that feature empirical data [16]. It provides an analytical method that is valid and robust because of the minimal requirements pertaining to sample size and the independence of data distributions [14].

Due to time and cost constraints, we surveyed 187 participants in a single-semester course. Therefore, PLS-SEM’s tolerance of small sample sizes and non-normal data were major advantages for our research. PLS-SEM emphasizes prediction likelihood and assesses dependent and independent variables in two steps. This creates a measurement model (also referred to as the outer model) and a structural model (also referred to as the inner model) [16].

III. RESEARCH MODEL AND HYPOTHESIS DEVELOPMENT

A. CONSTRUCTS AND MODERATING FACTORS

Figure 1 represents the conceptual framework of the UTAUT2 model, including demographic variables that influence technology acceptance. Within this framework, students’ behavioral intentions are considered predictors of future use [23].

![Conceputal framework of the UTAUT2 model](image)

FIGURE 1. Conceptual framework of the UTAUT2 model [13].

We opted to exclude the additional latent variables of ‘price value’ and ‘habit’ because they were not relevant to our research context. (The OJ system is provided free for educational use, and students will not make daily use of the system.) In addition, we excluded the dependent variable ‘use behavior’. This variable reflects the user behavior exhibited after using the technology [24], [25]. As the OJ system is made available as part of a course and is not a consumer product, this construct was also not relevant to our research context.

We further included the moderating variable of ‘academic major’ (MJR) to investigate whether different academic backgrounds influence students’ behavioral intentions. Participants in this study were grouped as follows: students from the Department of Information Engineering and Computer Science (IECS), students from the Department of Electrical Engineering (EE), and students from other departments. IECS students had been previously enrolled in other programming courses and likely considered programming to be an important professional skill for their major. EE students had taken at least one previous programming course and should exhibit a moderate level of interest in this skill. Moreover, students’ backgrounds, motivations, and attitudes are likely to influence their perceptions of the OJ system.

Earlier studies that applied the UTAUT2 set a precedent for adapting the model to suit their research aims. For example, Abu and Aljaafreh [26] identified the factors that affect students’ usage of social networking sites for educational purposes in Jordanian universities. To achieve this goal, they extended the UTAUT2 by including additional external factors, such as lecture support and student-related factors. Gansser and Reich [27] investigated the influencing factors in an acceptance model for products that employ artificial intelligence in everyday environments. Additional variables included personal innovativeness, security, health, convenience, and sustainability.

B. HYPOTHESES

1. Performance expectancy (PE)

‘Performance expectancy’ (PE) is defined as the degree to which an individual believes that using a system will enhance his or her performance. Almaih and Alyoussef [25] found that PE positively impacts the acceptance of e-learning systems. Cheng et al. [28] demonstrated that PE has an influential effect on the continuous use of news apps. Frank and Milkovic [29] found a significant correlation between PE and the adoption of electronic program guides.

The OJ system provides instant feedback. To make a submission, the code is zipped (compressed) and uploaded to the learning management system. The OJ then downloads, runs, and checks the code in the running environment. Therefore, students have the opportunity to improve their code quality, as their programs must pass all test cases. This process supports the development of students’ coding skills, which in turn builds confidence and satisfaction with the OJ system. Accordingly, we formulated the following hypothesis:

$H1$: Performance expectancy positively affects behavioral intention to use an online judge system in a programming course.

2. Effort expectancy (EE)

‘Effort expectancy’ (EE) is an important independent variable in our model. The main goal of including EE is to understand the degree of ease associated with the OJ
system. Almaiah and Alyoussef [25] determined that EE is a significant variable affecting the use of e-learning systems. Cheng [30] demonstrated a positive correlation between EE and the intention to use remote robotics experimentation in programming courses. Under a traditional approach, students write, compile, and test their code in their own environment. Additional efforts are required to become familiar with the operation of the OJ system. Integration of the OJ system with the OpenEdu system through a single sign-on methodology may confuse some students. The OJ system also strictly checks test cases, which may further confuse some students.

Structurally, the OJ system is simple for use. Students can easily access the assignments and their descriptions, sample inputs, expected outputs, and coding areas to write their code. After submitting the code, students can see the submission status (correct, compiler error, run time error, or wrong answer), as well as how long they took to complete each task. It seems likely that the students’ level of comfort in operating this system would be related to their past experience with programming environments. Thus, students registered for engineering and computer science degrees might expend less effort when using this technology. Therefore, we examined the moderating effect of ‘academic major’ (MJR) on the relationship between EE and intention to use the OJ system. This gives rise to the second and third hypotheses:

H2: Effort expectancy positively affects behavioral intention to use an online judge system in a programming course.

H3: Academic majors moderate the causal relationship between effort expectancy and behavioral intention to use the online judge system in a programming course.

3. Social influence (SI)
The construct of ‘social influence’ (SI) highlights the significant role that other people play in an individual’s decision to accept a certain technology. Jakaew and Hemrungrote [18] demonstrated that SI significantly affects the use of Google classrooms, while Meyliana et al. [31] confirmed the influence of SI on the use of learning management systems. Social support for technology acceptance can take many forms. For example, the IECS department at Feng Chia University built its own OJ system and encouraged senior students to help students with each other in this system. Therefore, we formulated the following hypothesis:

H4: Social influence positively affects behavioral intention to use an online judge system in a programming course.

4. Facilitating conditions (FC)
‘Facilitating conditions’ (FC) refer to the resources and support provided by an organization to aid in the adoption of a new system. Almaiah and Alyoussef [25] demonstrated that FC is an influential element in the use of e-learning systems, while Cheng [30] found a positive correlation between FC and the intention to use remote robotics experimentation in programming courses. The OJ system verifies the code by comparing the output of the code with the expected result, whereby even the smallest deviation will result in failure. For example, if the result is "12" but the student's code output is "12" the test will fail. This kind of error can be difficult to identify, which can result in frustration for students. Thus, guidance is especially important in such cases. If students do not possess the necessary skills, they may not accept the OJ system. Therefore, students' access to other resources may affect their behavioral intention to use the OJ system. Accordingly, we formulated the following hypothesis:

H5: Facilitating conditions positively affect behavioral intention to use the online judge system in a programming course.

5. Hedonic motivation (HM)
To achieve our research aims, we added the variable of ‘hedonic motivation’ (HM) to our research framework. HM represents the pleasure, enjoyment, and entertainment that an individual experiences when using a new technology. HM has been found to be a key factor influencing the acceptance and use of technology. For example, Salloum et al. [20] found that perceived enjoyment was the most important factor influencing the adoption of e-learning systems. Paulo and José Carlos [32] also applied PLS-SEM to measure the effects of HM on mobile health adoption and found that it had strong predictive power.

It is important to motivate students to regularly revisit course materials, especially in the early stages of distance learning. Interest is necessary to create positive emotions that can be used to measure student satisfaction. The OJ system examined in this study included a gamified component, and five of its six units included gamified programming exercises. These short exercises keep students bored as they progress through the material. Students must successfully pass their current units before they can progress. Thus, we propose the following hypothesis:

H6: Hedonic motivation positively affects behavioral intention to use the online judge system in a programming course.

It seems likely that previous experience working with programming platforms is likely to positively impact the enjoyment of the OJ system. Moreover, students who have registered for engineering and computer science degrees are likely to have a strong interest in programming. Therefore, we examined whether MJR moderated the relationship between HM and the intention to use the OJ system. This gives us the seventh hypothesis:

H7: Academic majors moderate the causal relationship between hedonic motivation and behavioral intention to use the online judge system in a programming course.

6. Self-efficacy (SE)
‘Self-efficacy’ (SE) refers to students’ sense of independence and/or self-confidence that enables them to problem-solve when something goes wrong. Almaiah et al. [24] found that SE has a significant effect on the acceptance
and use of mobile learning systems. Al-Rahmi et al. [33] further demonstrated that SE has a positive effect on the perceived usefulness of e-learning. Murked et al. [34] showed that SE positively affects the intention to adopt an electronic record management system in the educational sector. In a distance-learning environment, the ability to complete an exercise without intervention may strongly affect students’ behavioral intention to use the OJ system. Thus, we propose the following hypothesis:

**H8:** Self-efficacy positively affects behavioral intention to use an online judge system in a programming course.

Students may be more confident if they have previous experience using the OJ system. Students who are enrolled in other programming-related courses are likely to have more experience completing programming tasks and will have more opportunities to increase their self-efficacy. Therefore, we formulated our ninth hypothesis as follows:

**H9:** Academic majors moderate the causal relationship between self-efficacy and behavioral intention to use the online judge system in a programming course.

### 7. Anxiety (ANX)

‘Anxiety’ (ANX) is often experienced when using information technology, particularly new technology, and can be associated with the desire to avoid mistakes. Venkatesh et al. [12] revealed that computer ANX is a nonsignificant factor in the use and acceptance of information technology. The OJ system was designed to provide feedback by testing the submitted code. A previous study revealed that, on average, a student submitted 12 versions of a program before it passed. This design can make students feel anxious. Students may also need to develop confidence that the system will successfully store their work (e.g., that their work will not easily be lost if they hit the wrong button). Moreover, student rankings are always visible on the system, which may increase their anxiety about their final grades. These negative feelings can adversely impact the intention to use the OJ system. Thus, we formulated the following hypotheses:

**H10:** Anxiety negatively affects behavioral intention to use the online judge system in a programming course.

However, anxiety is likely to be influenced by experience and motivation. Therefore, we hypothesized that the effects of anxiety would be moderated by students’ choice of academic major.

**H11:** Academic majors moderate the causal relationship between anxiety and behavioral intention to use the online judge system in a programming course.

### 8. Behavioral intention (BI)

In this study, ‘behavioral intention’ (BI) is a dependent variable. Numerous studies have shown that BI mediates future use of technology. For example, Venkatesh et al. [12] confirmed that BI has a significantly positive influence on the use of information technology. In [25], BI was found to influence the use of e-learning systems.

Although students will not make everyday use of the OJ system after they complete the course, they can continue to use open systems such as Leetcode to enhance their programming practice. Students who intend to pursue a career in programming are more likely to continue using open-source OJ systems. Therefore, we hypothesized that academic MJR directly affects BI, as follows:

**H12:** Academic majors positively affect behavioral intention to use the online judge system in a programming course.

The 12 hypotheses were divided into two groups. The first group includes direct hypotheses pertaining to personalization variables (i.e., PE, EE, SI, FC, HM, SE, and ANX). The second group focused on the effects of academic MJRs. The relationships between the constructs are shown in Fig. 2.

### IV. RESEARCH METHODOLOGY

In this study, we employed a quantitative approach to evaluate the proposed model. This is in line with the approaches of previous studies that investigated technology use and acceptance.

#### A. MEASUREMENTS

The constructs of UTAUT2 were measured using the scales proposed in [35] and [36]. The additional variables SE and ANX were measured on scales adapted from [12], and HM was measured according to the suggestions made in [17]. We applied PLS-SEM software (version 3.0) to analyze the collected data. PLS is appropriate for early-stage research models, where the emphasis is on theoretical exploration and prediction. It does not require multivariate normality to estimate parameters, and it can be used with smaller sample sizes [37].

#### B. SURVEY DESIGN
Based on the research methods of the information technology system, all constructs were measured using a five-point Likert scale. We designed the items of each construct based on the UTAUT2 model with a total of 22 items, as shown in the table of the Appendix. A five-point Likert scale was used to measure the items that represented each construct, ranging from 1 (strongly disagree) to 5 (strongly agree). We published a survey and collected responses after the students finished the OJ exercises. Demographic questions were used to describe descriptive analysis and moderating effects.

C. COURSE DESCRIPTION

To recruit study participants, we selected a data science course offered by the College of Information and Electrical Engineering and was open to all FCU students. The course consisted of three parts: programming, machine learning, and applications. The students used the OJ system in the first six weeks, and our experiment was conducted in the first seven weeks. Nearly 300 undergraduate students were enrolled in this course, and 187 undergraduate students voluntarily responded to our survey.

This three-credit course included a combination of two-hour self-learning online and one-hour in-class learning per week. Before entering the physical classroom, students must watch videos, perform self-tests, and complete a series of OJ exercises. When learning online, students can discuss issues that arise with teaching assistants through a forum in the learning management system. In the physical class, teachers answer students’ questions; however, they do not repeat the lecture given in the video.

V. RESULTS

A. DESCRIPTIVE ANALYSIS

Table 1 summarizes the respondents’ demographic data. Most of the students who participated in our study were registered with the Department of Information Engineering and Computer Science (IECS) (n=156, 83.4%), followed by students from the Department of Electrical Engineering (EE) (n=23, 12.3%). Only eight students were from other departments (4.3%). Most of the respondents were male (80.2%), and the majority of respondents (38.5%) were in their third year of study, while 14.0% were in their fourth year of study. Finally, most respondents (80.2%) had previous experience using the OJ system. Only 19.8 had zero experience with this kind of system.

| TABLE 1. Demographic characteristics of study respondents (n=187). |
|-----------------|-----------------|-----------------|-------------------|
| Demographic Items | Categories          | Frequency | Percentage % |
| Major           | Department of IECS | 156       | 83.4%            |
|                 | Department of EE   | 23        | 12.3%            |
| Gender          | Male               | 150       | 80.2%            |
|                 | Female             | 37        | 19.8%            |
|                 | First year         | 36        | 19.2%            |
|                 | Second year        | 53        | 28.3%            |
|                 | Third year         | 72        | 38.5%            |

| Notes: IECS - Department of Information Engineering and Computer Science; EE - Department of Electrical Engineering |

B. ASSESSMENT OF MEASUREMENT MODEL

To calculate item reliability, internal consistency, and convergent validity (via confirmatory factor analysis), the measurement model was assessed using the PLS algorithm. When the PLS algorithm is employed, four reliability coefficients are typically applied: Cronbach’s coefficient alpha (α), composite reliability (CR), cross loadings (CL), and average variance extracted (AVE) [38]. If the indicators are highly correlated and interchangeable, they are considered reflective, and their reliability and validity should be thoroughly examined [39]. According to Ken [40], to be considered acceptable, ‘CL’, ‘α’, and ‘CR’ values should be equal to or greater than 0.7, and ‘AVE’ values should be greater than 0.5. The measurement results are presented in Table 2 and Fig. 3.

| TABLE 2. Internal consistency, item reliability, and convergent validity. |
|-----------------|-----------------|-----------------|-----------------|-----------------|
| Constructs      | Items           | CL (>0.7) | α (>0.7) | CR (>0.7) | AVE (>0.5) |
| Anxiety         | ANX1            | 0.956  | 0.753  | 0.879  | 0.786      |
|                 | ANX2            | 0.810  |         |         |            |
| Behavioral      | BI1             | 0.918  | 0.708  | 0.870  | 0.770      |
|                 | BI2             | 0.835  |         |         |            |
| Intention       | EE1             | 0.872  | 0.769  | 0.895  | 0.810      |
|                 | EE2             | 0.927  |         |         |            |
| Expectancy      | EE2             | 0.927  |         |         |            |
| Facilitating    | FC1             | 0.816  | 0.794  | 0.879  | 0.709      |
| Conditions      | FC2             | 0.877  |         |         |            |
|                 | FC3             | 0.831  |         |         |            |
| Hedonic         | HM1             | 0.965  | 0.924  | 0.963  | 0.929      |
| Motivation      | HM2             | 0.962  |         |         |            |
| Performance     | PE1             | 0.837  | 0.827  | 0.896  | 0.741      |
|                 | PE2             | 0.866  |         |         |            |
|                 | PE3             | 0.880  |         |         |            |
| Self-Efficacy   | SE1             | 0.940  | 0.867  | 0.938  | 0.882      |
|                 | SE2             | 0.938  |         |         |            |
| Social Influence| SI1             | 0.914  | 0.752  | 0.889  | 0.800      |
|                 | SI2             | 0.875  |         |         |            |
| MIR*ANX         | MIR*ANX1        | 0.994  | 0.783  | 0.859  | 0.757      |
|                 | MIR*ANX2        | 0.726  |         |         |            |
| MIR*EE          | MIR*EE1         | 0.779  | 0.746  | 0.870  | 0.771      |
|                 | MIR*EE2         | 0.968  |         |         |            |
| MIR*HM          | MIR*HM1         | 0.953  | 0.903  | 0.954  | 0.912      |
|                 | MIR*HM2         | 0.957  |         |         |            |
| MIR*SE          | MIR*SE1         | 0.978  | 0.892  | 0.944  | 0.893      |
|                 | MIR*SE2         | 0.911  |         |         |            |
| MIR->BI         | MIR*BI          | 1.000  |         |         |            |

| Notes: MIR - academic major; CL - cross loadings; α - Cronbach’s alpha; CR - composite reliability; AVE - average variance extracted. |

As shown in Table 2, all cross-loading values were above 0.726, and Cronbach’s alpha values were all above 0.708. This indicates that the reliability of the measurement model (outer model) was good. Furthermore, all composite reliability values exceeded 0.870, indicating good internal consistency. Finally, the average variance extracted was
above 0.709, which implies that the convergent validity of the proposed model is high.

The next important step in model verification is the Fornell-Larcker criterion analysis for discriminant validity. In this analysis, the results are considered acceptable when the values on the diagonal are higher than the cross-loading values of other constructs [40]. Good discriminant validity indicates that the constructs have sufficient convergent validity. The results of the Fornell-Larcker criterion analysis are shown in Tables 3 and 4.

### TABLE 3. Results of the Fornell-Larcker criterion (discriminant validity test-1).

|      | ANX | BI  | EE  | FC  | HM  | PE  | SE  | SI  |
|------|-----|-----|-----|-----|-----|-----|-----|-----|
| ANX  | 0.886|     |     |     |     |     |     |     |
| BI   | -0.141| 0.877|     |     |     |     |     |     |
| EE   | -0.258| 0.597| 0.900|     |     |     |     |     |
| FC   | -0.162| 0.763| 0.735| 0.842|     |     |     |     |
| HM   | -0.144| 0.795| 0.618| 0.807| 0.964|     |     |     |
| PE   | -0.168| 0.664| 0.676| 0.714| 0.732| 0.861|     |     |
| SE   | -0.154| 0.685| 0.508| 0.671| 0.548| 0.400| 0.939|     |
| SI   | -0.143| 0.749| 0.707| 0.820| 0.782| 0.794| 0.529| 0.895|

Notes: ANX - anxiety; BI - behavioral intention; EE - effort expectancy; FC - facilitating conditions; HM - hedonic motivation; PE - performance expectancy; SE - self-efficacy; SI - social influence.

### TABLE 4. Results of the Fornell-Larcker criterion (discriminant validity test-2).

|      | MJR*ANX | MJR*EE | MJR*HM | MJR*SE | MJR*BH |
|------|---------|--------|--------|--------|--------|
| ANX* | -0.870  | 0.878  |        |        |        |
| BI   | -0.033  | 0.878  |        |        |        |
| EE   | -0.108  | 0.474  | 0.955  |        |        |
| HM   | -0.133  | 0.299  | 0.688  | 0.945  |        |
| SE   | -0.073  | 0.138  | 0.173  | -0.051 |        |

Notes: ANX - anxiety; BI - behavioral intention; EE - effort expectancy; HM - hedonic motivation; SE - self-efficacy; MJR - academic major.

### C. ASSESSMENT OF STRUCTURAL MODEL

Next, we assessed how the latent variables relate to each other in the structural model. For this, the blindfolding technique was used to evaluate the predictive relevance of the proposed research model. Under this technique, if a cross-validated redundancy measure “Q²” is larger than zero, the model is considered to have predictive relevance. Moreover, when the explained variance “R²” value is greater than 0.67, it is perceived to be “substantial” [16], [40], as shown in Table 5.

### TABLE 5. Explained variance (R²) and predictive relevance (Q²)

| Dependent variables | R²(>0.5) | Result | Q²(>0) | Decision |
|---------------------|----------|--------|--------|----------|
| Behavioral Intention | 0.754    | High   | 0.532  | Model has predictive relevance |

For this analysis, the results of the p-value calculations were obtained via bootstrapping and two-tailed tests. Bootstrapping is a nonparametric technique that randomly resamples the original dataset to estimate the statistical significance of the PLS path model [39], [41]. The results are listed in Table 6.

### TABLE 6. Results of testing study hypotheses related to direct effects (p < 0.01, p < 0.05)

| H     | Relation | Original Sample (M) | Sample Mean (M) | Standard Deviation (SD) | T Statistics (1.96) | P Values |
|-------|----------|---------------------|-----------------|------------------------|---------------------|----------|
| H1    | PE -> BI | 0.066               | 0.066           | 0.074                  | 0.891               | 0.373    |
| H2    | EE -> BI | -0.035              | -0.019          | 0.076                  | 0.457               | 0.648    |
| H4    | SI -> BI | 0.245               | 0.250           | 0.102                  | 2.391               | 0.017*   |
| H5    | FC -> BI | 0.010               | -0.009          | 0.104                  | 0.098               | 0.922    |
| H6    | HM -> BI | 0.400               | 0.388           | 0.080                  | 5.008               | 0.000**  |
| H8    | SE -> BI | 0.316               | 0.323           | 0.067                  | 4.731               | 0.000**  |
| H10   | ANX -> BI | 0.011              | 0.005          | 0.038                  | 2.239               | 0.026*   |
| H12   | MJR -> BI | -0.098              | -0.093         | 0.044                  | 2.239               | 0.026*   |

Notes: ANX - anxiety; BI - behavioral intention; EE - effort expectancy; FC - facilitating conditions; HM - hedonic motivation; PE - performance expectancy; SE - self-efficacy; SI - social influence; MJR - academic major; H - hypothesis; S - supported; NS - non-supported

As shown in Table 6, four variables had significant predictive relationships with ‘behavioral intention’: ‘hedonic motivation’ (p=0.000, t=5.008), ‘self-efficacy’ (p=0.000, t=4.731), ‘social influence’ (p=0.017, t=2.391), and ‘academic major’ (p=0.026, t=2.239). In other words, H4, H6, H8, and H12 are supported. In contrast, four variables had non-significant relationships with behavioral intention: ‘anxiety’ (p=0.778, t=0.282), ‘effort expectancy’ (p=0.648, t=0.457), ‘facilitating conditions’ (p=0.922, t=0.098), and ‘performance expectancy’ (p=0.373, t=0.891). Therefore, H1, H2, and H5 are not supported. Moreover, H10 was not supported by the P value.
calculation. This result confirmed that ANX was a negative factor in the intention to use the OJ system.

D. MODERATING EFFECTS

Under the PLS-SEM approach, when there is theoretical support for multiple moderators, the researcher may consider analyzing the relationships to improve the interpretability of results [39]. In the current study, we examined the moderating effects of ‘academic major’ on the relationships between ‘behavioral intention’ (the dependent variable) and ‘anxiety’, ‘effort expectancy’, ‘hedonic motivation’, and ‘self-efficacy’ (independent variables). The results of the analysis are presented in Table 7.

| H    | Relation                | Original Sample (O) | Sample Mean (M) | SD   | T-Statistic (>1.96) | P Values (<=0.05) |
|------|-------------------------|---------------------|-----------------|------|--------------------|------------------|
| H3   | MJR*EE -> BI            | 0.064               | 0.060           | 0.065| 0.985              | 0.325            |
| H7   | MJR*HM -> BI            | 0.006               | -0.013          | 0.083| 0.067              | 0.947            |
| H9   | MJR*SE -> BI            | -0.002              | 0.016           | 0.076| 0.029              | 0.977            |
| H11  | MJR*ANX -> BI           | 0.060               | 0.053           | 0.065| 0.921              | 0.358            |

Notes: ANX - anxiety; EE - effort expectancy; HM - hedonic motivation; SE - self-efficacy; H - hypothesis; S - supported; NS - non-supported. MJR - academic major.

These results reveal that ‘academic major’ played a non-significant moderating role in relationships between ‘behavioral intention’ and the following variables: ‘anxiety’ (p=0.358, t=0.921), ‘effort expectancy’ (p=0.325, t=0.985), ‘hedonic motivation’ (p=0.947, t=0.667), and ‘self-efficacy’ (p=0.977, t=0.299). Therefore, H3, H7, H9, and H11 were not supported. In other words, academic majors did not exert a significant moderating effect on the relationships considered.

VI. DISCUSSION AND IMPLICATIONS

In this study, we applied UTAUT2 and PLS-SEM to evaluate the acceptance of the OJ system by students in an online programming course. Our findings add to the existing literature pertaining to the use of OJ systems in distance learning. This paper is the first to use the UTAUT2 model in conjunction with demographic characteristics to investigate the acceptance of an OJ system for online programming.

Our results can serve as a reference for the development of OJ systems for computer science education at the university level. Results pertaining to ‘hedonic motivation’ (i.e., p=0.000, t=5.008) imply that universities and OJ system developers need to focus on ongoing motivation to help students successfully complete online courses. This can be achieved by gamifying the system further. It may also be worth exploring ways to reduce the anxiety that arises when students use the system.

Our results pertaining to ‘self-efficacy’ (p=0.000, t=4.731) show that self-confidence helps students when they encounter mistakes or difficulties during their programming practice. This positive effect motivates them to continue learning in the face of challenges. Self-efficacy is a learning habit that can be cultivated by students. It is worth noting the effects of ‘social influence’ (p=0.017, t=2.391), which is related to ‘self-efficacy’. The success of distance learning is influenced by relationships with friends, classmates, family, and teachers. The relationship between ‘social influence’ and ‘self-efficacy’ highlights the importance of positive social support for students. Positive social support can be facilitated by instructors who stay in touch with their students through a variety of communication tools, such as discussion forums, e-mail, and messaging systems. In other words, supportive instructors typically have a positive effect on students’ final grades and achievements.

Interestingly, while ‘academic majors’ exerted only nonsignificant effects on the other relationships examined in this study, it nonetheless exhibited a significant, direct influence on behavioral intention (p=0.026, t=2.239). Taken together, our findings suggest that if students are strongly self-motivated and/or highly self-confident, their academic major will not have a strong impact on their acceptance of the OJ system. In other words, students can successfully complete an online programming course with the OJ system when they are able to study independently, are confident, and feel supported by teachers and others.

It is also worth noting that ‘anxiety’ had a non-significant effect on students’ behavioral intentions (p=0.778, t=0.282). High levels of student anxiety can greatly negatively affect their desire to learn, their self-confidence when faced with a problem, and especially their ability to successfully complete distance learning. Programming exercises often require students to make more decision-making processes. Therefore, university, teachers, and OJ system developers should pay attention to this issue.

VII. LIMITATIONS AND DIRECTIONS FOR FUTURE WORK

Although the results of this study provide valuable contributions, certain limitations must be noted. First, the study respondents only included undergraduate students from a single university course. Therefore, we recommend that future research include a wider range of participants. For example, a future study could recruit participants from massive open online courses. Second, our results indicated that the demographic characteristics of academic majors did not exert strong effects. Moreover, as the participants in our study were similar in age, we could not examine the effects of the age moderating variable, which is one of the basic moderations. Therefore, we suggest that future research investigate this important factor in the context of UTAUT2 if there is a significant age difference between study participants.
VIII. CONCLUSIONS
The OJ is widely used in online programming courses; however, no previous research has evaluated its acceptance by students. Furthermore, while many studies have focused on improving the OJ system, UTAUT2 variables have only been applied in the context of mobile apps, home technologies, businesses, banks, e-learning, and information technology.

We found that the most influential factors in the proposed model were ‘hedonic motivation’, ‘self-efficacy’, and ‘social influence’. Interestingly, academic majors did not play a significant moderating role between the intention to use the OJ system and other influential factors. However, it exerted a significant direct effect on ‘behavioral intention’. In summary, students can successfully complete OJ-based online programming courses regardless of their academic majors when they are independent, self-confident, and supported by teachers and others.

APPENDIX

TABLE. Measurement items in estimating UTAUT2 model [12], [13], [17], [35].

| Items | Descriptions |
|-------|--------------|
| PE1   | I think using the OJ system enables me to know earlier if any part of my code is correct or not. |
| PE2   | I think using the OJ system to submit HW in programming courses is better than learning traditional methods. |
| PE3   | I think using the OJ system could increase programming quality. |
| EE1   | It would not take me long to learn how to use the OJ system. |
| EE2   | I think the OJ system would be easy to understand how to use. |
| SI1   | People who are important to me think that I should use the OJ system to enhance programming skills in the self-learning context. |
| SI2   | In general, the department has supported the use of the OJ system. |
| FC1   | I have the knowledge necessary to use the OJ system. |
| FC2   | The OJ system is compatible with other forms of programming exercises that I use. |
| FC3   | I can get help from others when I have difficulties using the OJ system. |
| HM1   | Practicing programming on the OJ system is enjoyable (challenging and interesting). |
| HM2   | Practicing programming on an OJ system gives me the motivation to continue learning. |
| SE1   | When something goes wrong, if there was no one around to tell me what to do, I can use the OJ system to do my homework or assignments without the help of others. |
| SE2   | When something goes wrong, if I had a lot of time to complete the job for which the software was provided, I can use the OJ system to do my homework or assignments without the help of others. |
| ANX1  | I feel apprehensive about using the OJ system. |
| ANX2  | It scares me to think that I could lose a lot of information using the system by hitting the wrong button. |
| B1    | I will take other courses that are using the OJ system. |
| B2    | I intend to continue using the OJ-like system (such as LeetCode) to enrich my career record when practicing programming in the future. |

ACKNOWLEDGMENT
This research was supported by the Ministry of Science and Technology, Taiwan, under grant MOST 109-2221-E-035-056.

REFERENCES
[1] R. Conejo, B. Barros, and M. F. Bertoa, "Automated Assessment of Complex Programming Tasks Using SIETTE," IEEE Transactions on Learning Technologies, vol. 12, no. 4, pp. 470-484, 2019.
[2] H.-M. Chen, B.-A. Nguyen, Y.-X. Yan, and C.-R. Dow, "Analysis of Learning Behavior in an Automated Programming Assessment Environment: A Code Quality Perspective," IEEE Access, vol. 8, pp. 167341-167354, 2020.
[3] I. Meketerovic, L. Brkic, B. Milasinovic, and M. Baranovic, "Building a Comprehensive Automated Programming Assessment System," IEEE Access, vol. 8, pp. 81154-81172, 2020.
[4] R. Sukamto, R. Megasari, E. Piantari, and M. Rischa, "Code Comment Assessment Development for Basic Programming Subject using Online Judge," presented at the Proceedings of the Proceedings of the 7th Mathematics, Science, and Computer Science Education International Seminar, MSCEIS 2019, 12 October 2019, Bandung, West Java, Indonesia, 2020.
[5] M. Wang, W. Han, and W. Chen, "MetaOJ: A Massive Distributed Online Judge System," Tsinghua Science and Technology, vol. 26, pp. 548–557, 2021.
[6] J. Arifin and R. S. Perdana, "UGrade: Autograder for Competitive Programming Using Contestant PC as Worker," presented at the International Conference on Data and Software Engineering (ICOSE), 2019.
[7] X. Lu, D. Zheng, and L. Liu, "Data Driven Analysis on the Effect of Online Judge System," presented at the International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData), 2017.
[8] F. Pereira et al., "Using learning analytics in the Amazonas: understanding students’ behaviour in introductory programming," British Journal of Educational Technology, vol. 51, no. 4, pp. 955-972, 2020.
[9] J. Petit et al., "Jutge.org: Characteristics and Experiences," IEEE Transactions on Learning Technologies, vol. 11, no. 3, pp. 321-333, 2018.
[10] J. L. Fernandez Aleman, "Automated Assessment in a Programming Tools Course," IEEE Transactions on Education, vol. 54, no. 4, pp. 576-581, 2011.
[11] D. Galan, R. Heradio, H. Vargas, I. Abad, and J. A. Cerrada, "Automated Assessment of Computer Programming Practices: The 8-Years UNED Experience," IEEE Access, vol. 7, pp. 130113-130119, 2019.
[12] V. Venkatesh, M. G. Morris, G. B. Davis, and F. D. Davis, "User Acceptance of Information Technology Toward a Unified View," MIS Quarterly, vol. 27, pp. 425-478, 2003.
[13] V. Venkatesh, J. Y. L. Thong, and X. Xu, "Consumer Acceptance and Use of Information technology: Extending the Unified Theory of Acceptance and Use of Technology," MIS Quarterly, vol. 36, pp. 157-178, 2012.
[14] H. M. Lin, M. H. Lee, J. C. Liang, H. Y. Chang, P. Huang, and C. C. Tsai, "A review of using partial least square structural equation modeling in e-learning research," British Journal of Educational Technology, vol. 51, no. 4, pp. 1354-1372, 2019.
[15] J. F. Hair, G. T. M. Hult, C. M. Ringle, and K. O. Thiele, "Mirror, mirror on the wall: a comparative evaluation of composite-based structural equation modeling methods," Journal of the Academy of Marketing Science, vol. 45, no. 5, pp. 616-632, 2017.
[16] J. F. Hair, C. M. Ringle, and M. Sarstedt, "PLS-SEM: Indeed, a Silver Bullet," Journal of Marketing Theory and Practice, vol. 19, no. 2, pp. 139-152, 2014.
[17] F.-H. Huang, "Adapting UTAUT2 to assess user acceptance of an escooter virtual reality service," Virtual Reality, vol. 24, no. 4, pp. 635-643, 2020.
[18] P. Jakkaw and S. Hemrngrote, "The Use of UTAUT2 Model for Understanding Student Perceptions Using Google Classroom: A Case Study of Introduction to Information Technology Course," IEEE Access, 2017.
[19] J. Y. L. Thong, V. Venkatesh, X. Xu, S.-J. Hong, and K. Y. Tam, "Consumer Acceptance of Personal Information and Communication Technology Services," IEEE Transactions on Engineering Management, vol. 58, no. 4, pp. 613-625, 2011.
This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3126896, IEEE Access