Journal of the Association for Information Systems

Volume 21 | Issue 4 Article 8

7-15-2020

Living in a Simulation? An Empirical Investigation of a Smart Driving-Simulation Testing System

Wen Xie
Hong Kong Polytechnic University, wen.xie@connect.polyu.hk

Xin Xu
Hong Kong Polytechnic University, xin.xu@polyu.edu.hk

Ruiqi Liu
Hong Kong Polytechnic University, ruiqi.liu@connect.polyu.hk

Yong Jin
Hong Kong Polytechnic University, jimmy.jin@polyu.edu.hk

Wenchao Bai
Hong Kong Polytechnic University, barrybwz@aliyun.com

See next page for additional authors

Recommended Citation
Xie, Wen; Xu, Xin; Liu, Ruiqi; Jin, Yong; Bai, Wenchao; and Li, Qiang (2020) "Living in a Simulation? An Empirical Investigation of a Smart Driving-Simulation Testing System," *Journal of the Association for Information Systems*, 21(4), .
DOI: 10.17705/1jais.00622
Available at: https://aisel.aisnet.org/jais/vol21/iss4/8

Journal of the Association for Information Systems (2020) **21**(4), 843-863
doi: 10.17705/1jais.00622

RESEARCH ARTICLE

ISSN 1536-9323

Living in a Simulation? An Empirical Investigation of a Smart Driving-Simulation Testing System

Wen Xie[1], Xin Xu[2], Ruiqi Liu[3], Yong Jin[4], Wenchao Bai[5], Qiang Li[6]
[1]Hong Kong Polytechnic University, Hong Kong, wen.xie@connect.polyu.hk
[2]Hong Kong Polytechnic University, Hong Kong, xin.xu@polyu.edu.hk
[3]Hong Kong Polytechnic University, Hong Kong, ruiqi.liu@connect.polyu.hk
[4]Hong Kong Polytechnic University, Hong Kong, jimmy.jin@polyu.edu.hk
[5]Hong Kong Polytechnic University, Hong Kong, barrybwz@aliyun.com
[6]South China Agricultural University, China, lq1128@vip.163.com

Abstract

The internet of things (IoT) generally refers to the embedding of computing and communication devices in various types of physical objects (e.g., automobiles) used in people's daily lives. This paper draws on feedback intervention theory to investigate the impact of IoT-enabled immediate feedback interventions on individual task performance. Our research context is a smart test-simulation service based on internet-of-vehicles (IoV) technology that was implemented by a large driver-training service provider in China. This system captures and analyzes data streams from onboard sensors and cameras installed in vehicles in real time and immediately provides individual students with information about errors made during simulation tests. We postulate that the focal smart service functions as a feedback intervention (FI) that can improve task performance. We also hypothesize that student training schedules moderate this effect and propose an interaction effect on student performance based on feedback timing and the number of FI cues. We collected data about students' demographics, their training session records, and information about their simulation test(s) and/or their official driving skills field tests and used a quasi-experimental method along with propensity score matching to empirically validate our research model. Difference-in-difference analysis and multiple regression results support the significant impact of the simulation test as an FI on student performance on the official driving skills field test. Our results also supported the interaction effect between feedback timing and the number of corrective FI cues on official test performance. This paper concludes with a discussion of the theoretical contributions and practical significance of our research.

Keywords: Internet of Things, Internet of Vehicles, Feedback Interventions, Feedback Timing, Quasi-Experiments, Driver Training.

Kenny Cheng and T.P. Liang were the accepting senior editors. This research article was submitted on July 29, 2018 and underwent two revisions.

1 Introduction

The internet of things (IoT) generally refers to the embedding of computing and communication devices into various types of physical objects (e.g., automobiles) used in people's daily lives in order to enable real-time data transmission between human beings and these devices over the internet (Wortmann & Flüchter 2015). IoT technologies enable real-time capturing, tracking, and processing of data about individual behaviors—i.e., the digitized/quantified selves (Rivera-Pelayo et al. 2012). Results from analysis of the massive amounts of data collected by IoT technologies can be used to enable a wide array of smart services (e.g. smart home, smart transportation, smart cities, etc.; Wortmann & Fluchter, 2015). McKinsey estimated the total market value of IoT applications at USD 900 million in 2015 and predicted that it will reach 3.7 billion by 2020 with a compound annual growth rate of 32.6%.[1] Global investment in IoT-based smart services is on the rise, with smart homes, smart wearables, and smart cities topping the list.[2] Given the huge market potential of smart services, businesses have a pressing need to understand how to fully tap into the IoT data streams to create high-value service innovations (Porter & Heppelmann, 2015).

Our research investigates the impacts of IoT smart services on individual users, focusing particularly on internet of vehicles (IoV) technology. IoV plays an important role in the entire IoT value system (ranked as the sixth most popular IoT topic on the web and the fourth most popular smart city IoT project; see Footnotes 1 and 2 below). The IoV infrastructure consists of devices connected over in-vehicle networks, intervehicle networks, and car-mounted mobile internet applications. Communication protocols and data exchange standards support huge volumes of wireless communication and information exchange between IoV-equipped cars and other cars, roads, pedestrians, the internet, and so forth, based on which smart services are implemented—for example, intelligent traffic management, smart information services, smart vehicle control, etc. IoV and the IoV data already generated have a number of potential commercial applications. For example, in the context of the insurance industry, by tracking and analyzing IoV data about drivers' microdriving behaviors (e.g., sharp turns and sudden braking), an insurance company would be able to more accurately determine the risk profile of each driver (e.g., the probability of traffic violations or odds of accidents). Car insurance premiums could then be optimized, thus helping to increase the market share or profitability of insurance companies (Soleymanian, Weinberg, & Zhu, 2016). The current study, however, focuses on an IoV smart service in the context of driver training in China

1.1 The Research Context: A Smart Driving-Simulation Test in China

The smart service we examine here is an IoV-based driving-simulation testing system adopted by a large driver-training service provider in China. In particular, the system is built to simulate the "Subject 2" (or K2) driving skills field test, which is administered to driver's license applicants in China. K2 tests various driving skills—such as including backing up and parking a car, parking and starting a car on a hill, right- and left-hand turns, changing lanes and passing cars, parallel parking, etc. During the official K2 test, an examiner accompanies the applicant while they complete all the K2 tasks and then scores the applicant's driving proficiency. Typically, before taking the test, individuals enroll in a training program offered by an accredited driving school that pairs students with a driving instructor who teaches them basic driving skills over the course of the required training hours (16 hours for K2). After that, the individual takes the official K2 test. Individuals scoring over 90% pass; those that fail the test can retake it at a later date. Individuals who successfully pass the K2 test are then eligible to prepare and take the Subject 3 (K3) driving skills road test.

The IoV-based driving-simulation testing system targets the K2 test only. It was adopted by the focal driving school in order to improve student performance on the actual test. The simulation system mimics an actual field test and vehicles are equipped with internet-enabled sensors and in-vehicle cameras. During the simulation test, the driving student performs various driving tasks, just as they would during the official test. The system automatically captures, tracks, and analyzes real-time data streams about the driving behaviors of individual students and provides feedback on the results of their driving tests directly following the simulation test (see a sample feedback report in Appendix A). From the student perspective, the system is smart because (1) it can detect and capture every error a driving student makes during the simulation test while performing driving tasks in an authentic field setting, and (2) it generates a real-time report immediately following the simulation test complete with the total score, details about the errors made, and photos of the student's driving actions performed during the simulation test. From the driving school's perspective, this smart service reduces operating costs by eliminating the need to provide human examiners during the simulation test and also adds value to the training program overall. Figure 1 depicts the timeline of a typical K2 training program, the simulation test, and the official test. It also shows the metrics of the key variables in our study—Feedback Timing, Performance Metric 1, and Performance Metric 2 (see our methods section for details).

In this paper, we first examine whether the adoption of the IoV-based simulation testing system improves performance on the official driving test. We then investigate the role of the training schedule (e.g., the mean and the standard deviation of training session intervals) as the boundary condition of the simulation test effect. Finally, we evaluate whether the timing of the simulation test affects performance using the simulation results as the moderator.

[1] https://iot-analytics.com/10-internet-of-things-applications/

[2] https://iot-analytics.com/top-10-iot-segments-2018-real-iot-projects/

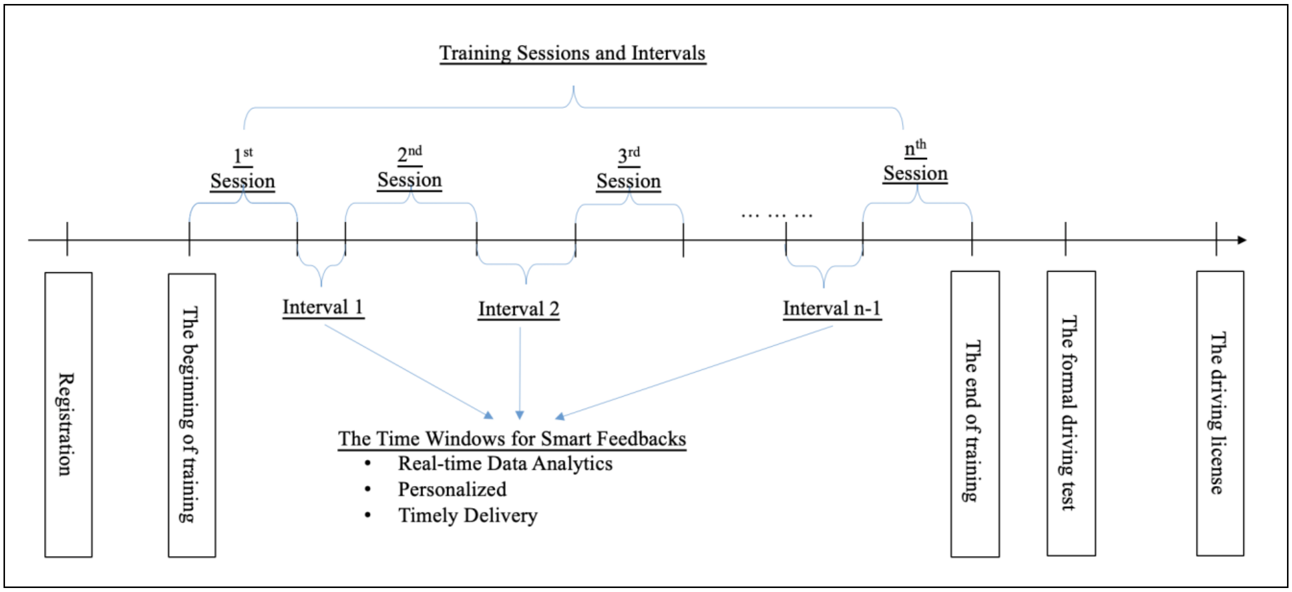
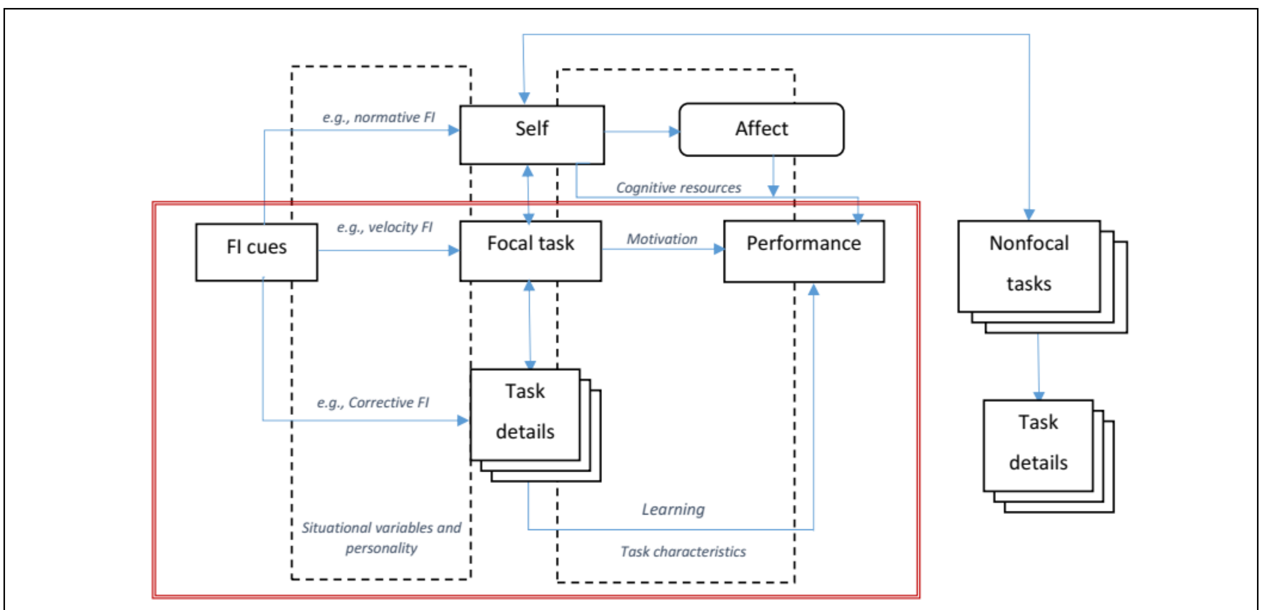


Figure 1. The Timeline and Key Metrics



Note: We mainly use the part within the red box in Figure 2 as our theoretical foundation in the current paper.

Figure 2. Feedback Intervention Theory

1.2 Literature and Theoretical Foundation

Research on IoV-based smart services is limited in major business and transportation research journals. We conducted a comprehensive review and found an in-depth understanding of the user impacts of smart services to be lacking (Lee & Lee, 2015). In particular, we found that most of the current literature focuses on technology evolution and prospective business applications (e.g., Lee & Lee, 2015; Wortmann & Flüchter, 2015). Both an overarching theoretical foundation and specific research models are needed to better understand the impacts of IoT services on users. Although recent design science research has proposed a few conceptual frameworks based on reflective learning theory (e.g., Müller, Rivera-Pelayo, & Heuer, 2012; Rivera-Pelayo et al., 2012), the frameworks are still too broad to produce concrete propositions/hypotheses that can be tested by empirical studies.

Furthermore, we found that existing empirical studies of driver behaviors are mainly based on data from onboard devices (OBDs) that are typically collected in batch mode. For example, transportation research has primarily focused on predicting driving risk/safety using OBD data and has found that specific driving behaviors (e.g., speeding, sudden braking, etc.) can better predict individuals' driving risks (Paefgen et al., 2014). The value of timely information derived from analyzing IoV data streams has not yet been investigated.

To address these gaps in the literature, this paper adopts feedback intervention theory (Kluger & DeNisi, 1996; see Figure 2) as the theoretical foundation and overarching framework of our research. Feedback intervention theory (FIT) is rooted in control theory, which posits that people change their behaviors when they are motivated by a performance gap between their current behavior and a goal. FIT suggests that such goals are hierarchically ordered on three levels—namely, the self level, the focal-task level, and the task-detail level. Different feedback interventions (FIs) direct individuals' attention toward different goal levels, which, in turn, influence behavioral performance through different mechanisms. For instance, velocity FIs concerning overall performance/progress impact the focal-task goal, which, in turn, affects overall task performance via the *motivation mechanism*—e.g., through an individual's desire to fill the performance gap. In contrast, corrective FIs concerning behavioral errors motivate attention to task details, which improves performance through the learning mechanism—i.e., by mastering the skills needed to successfully perform the focal task. Finally, situational variables and task characteristics are the boundary conditions that allow FI cues to impact goals and, in the end, performance.

We believe that FIT provides a solid theoretical foundation for studying the impacts of information derived from IoT data streams on individual users. The huge volume of IoT data on both physical objects and human behaviors is comprised of microdata that are aggregated to produce information useful for actual human applications. For instance, in our research context, IoV, microdata concerning the basic features of a car (e.g., brakes, speed, etc.) and human actions (e.g., head movements, directions of eyesight, etc.) can be aggregated to reflect a driver's proficiency at the task-detail level—e.g., in terms of parking skill. Aggregating task details, in turn, reflects overall task performance—e.g., overall level of driving skills and, thus, probability of passing the driving skills field test. FIT's specification of the three goal levels corresponds to the different levels of information that can be generated by aggregating IoT data streams and communicating their significance to users. Furthermore, FIT also indicates how IoT data streams may be used to create different FI cues aimed at improving performance through various mechanisms. For instance, in our research context, FIT suggests that information about parking errors would direct a driver's attention toward understanding the correct steps needed to park a car, which, in turn, would improve performance the next time he or she attempted to park a car via the learning mechanism.

In summary, there is a lack of theoretical development and empirical study of IoT-based smart services in the existing research. Our paper aims to fill this gap by adopting FIT as the theoretical foundation in the context of an IoV-based driving-simulation test in order to generate insights into how smart services based on IoV data streams impact individual users. We present two research questions: (1) *Will an IoV-based simulation test significantly improve the official field test performance of driving license applicants? (2) Will the timing of the IoV-based simulation test and the number of feedback cues from the simulation jointly influence official field test performance?*

2 Hypotheses Development

As illustrated in Figure 3a, we first examined the impact of the IoV-based driving-simulation test on official driving skills field test performance and then investigate the effect of associated boundary conditions—i.e., the interaction effect of the simulation test and the training schedule on official test performance.

We predict that experience with the IoV-based simulation test will have a positive impact on official driving skills test performance. As discussed in the Introduction, IoV technology enables the simulation testing system to capture and analyze data streams generated from the sensors and the cameras installed in vehicles to produce and report simulation test results in real time. The results comprise both the total score a student receives on the simulation test and details of the errors made by the student during the simulation (see Appendix A). The results constitute the FI cues that we predict will impact on the student's later performance on the official driving skills test. In particular, the total score earned on a driving-simulation test operates as a velocity FI cue about overall performance, which will activate the *motivation mechanism* by directing the student's attention to the goal of the focal task—i.e., passing the official driving skills test. In contrast, information about errors made during the driving-simulation test serves as the corrective FI cue that will direct the driving student's attention toward task details (e.g., the steps necessary to correctly park a car), which will improve official test performance through the *learning mechanism*—i.e., through mastering the skills needed to pass the official test. In general, we predict that the immediate feedback cues provided by the simulation testing system will trigger both motivation and learning mechanisms that will improve performance on the official driving skills test.

H1: Driving students who participate in the simulation test will perform better on the official driving skills field test than those who do not.

We next propose that the simulation test will be more beneficial for driving students with a relatively regular training schedule (e.g., a small standard deviation of intervals between training sessions) than it will be for others. FIT suggests that there are boundary conditions such as situational variables or individual traits that moderate the impacts of FI cues on performance (Kluger & DeNisi, 1996). We focus on the moderating role of training regularity in the current study. Existing literature on learning suggests that regular learning or training enhances both mental and muscle memory and facilitates familiarity with knowledge and skills (Fleishman, 1972). In our research context, training regularity refers to the frequency with which driving students participate in training in order to acquire the knowledge and skills necessary to develop driving proficiency. The impact of corrective FI cues on learning task details depends on the individual's prior knowledge and experience related to the focal task (Kluger & DeNisi, 1996). We anticipate that, compared with others, driving students who participate in regular training sessions will become more quickly familiar with the knowledge and skills needed to drive, will have more opportunity to improve driving skills, will be able to more efficiently correct errors identified by the simulation testing system, and will therefore have a lower probability of making errors during the official driving skills test. In short, we predict that regular training sessions will enhance the positive impact of the IoV-based simulation test on official test performance.

H2: Regular training will positively moderate the impact of the simulation test on official test performance.

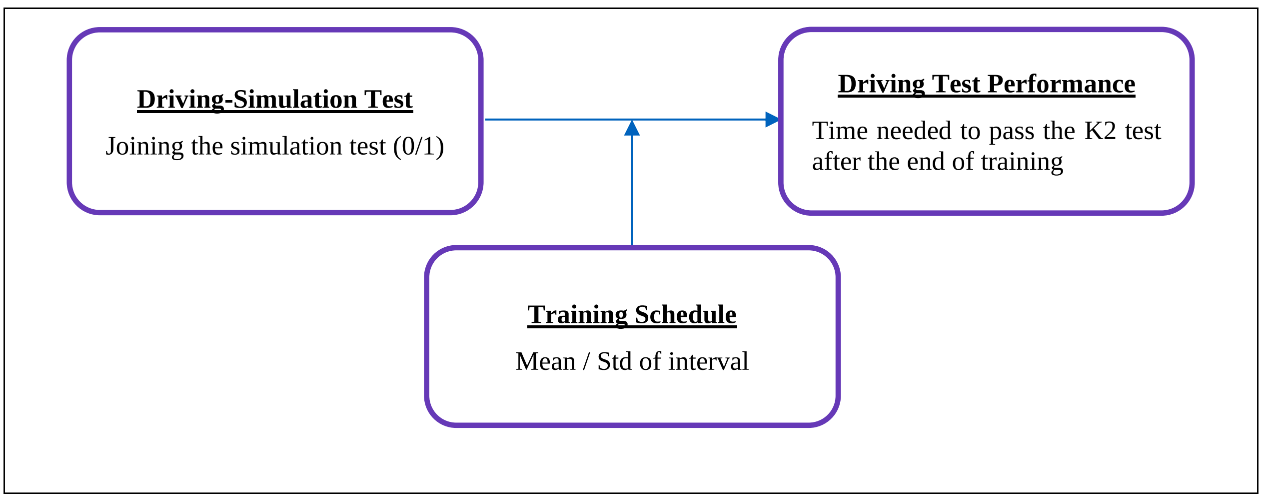


Figure 3a. Research Model 1: Simulation Test, Training Schedule, and Test Performance

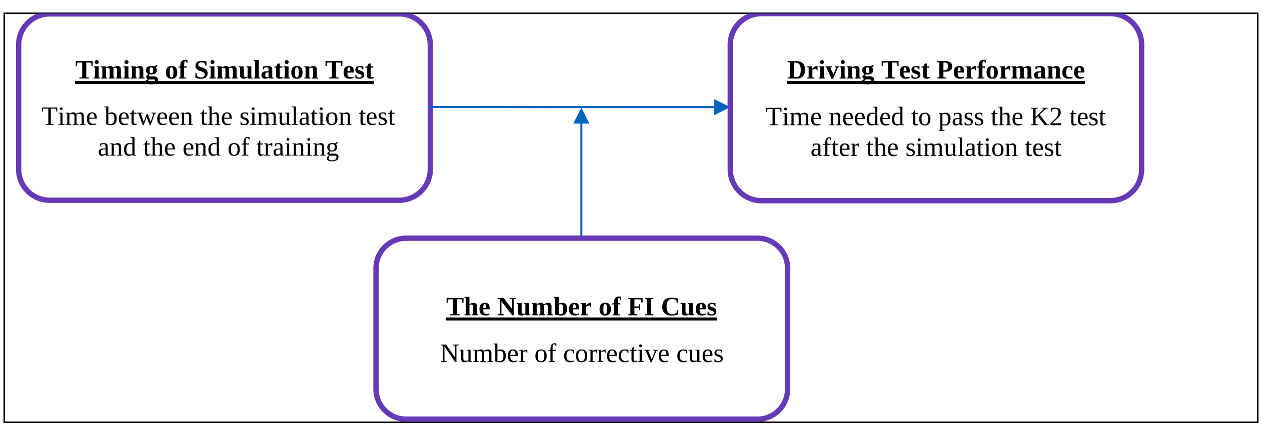


Figure 3b. Research Model 2: Feedback Timing, Number of Cues, and Performance

As shown in Figure 3b, we further examined the effect of the *timing of the simulation test* (i.e., feedback timing) on official test performance. The impact of feedback timing on performance has been extensively examined in the literature on learning and has produced mixed results (e.g., Butler, Karpicke, & Roediger, 2007). Overall, Kluger and DeNisi (1996) did not include an explicit notion of the role of feedback timing in FIT, probably because early literature on feedback did not emphasize feedback timing. However, FIT has identified the significance of task learning (in terms of task details and task characteristics), which is important for our understanding of the effect of feedback timing on task performance. Feedback timing influences task performance mainly by affecting how individuals absorb and memorize new knowledge and skills (i.e., the mechanism of task-detail learning in FIT). Furthermore, driving is a complex task that demands significant physical and mental effort (i.e., it is related to the task-complexity notion in FIT). We thus integrate recent literature on feedback timing, particularly concerning feedback timing in complex tasks, with FIT serving as the theoretical basis of our Hypothesis 3 below. In particular, we postulate that feedback timing influences task performance by affecting how details of complex tasks are mastered through the absorption of knowledge and skills and the retention of the task details in memory. Horneck (2016) examined feedback timing in the context of complex, multistep tasks (e.g., the maze game). He found that when performance feedback was given too early or too late, subjects had relatively poor task performance. In general, if feedback is given too early, subjects will not have enough time to absorb the knowledge through self-reflection. However, if feedback is given too late, subjects may forget what they learned. In either case, subjects will experience increased learning costs and/or cognitive load when processing feedback, potentially contributing to poor task performance.

In our research context, learning how to drive is a complex task demanding both mental and physical effort; it requires the completion of a multisession training program, which may include a final simulation test. While the length of the training period may vary across individual driving students, the timing of the simulation test is exogenously determined by the system, based on the date the student is scheduled to take the official driving skills field test. While the general rationale about feedback timing in Thornock (2016) may also apply to our context, we predict that a higher number of days between the end of training and the simulation test will exert a positive effect on the official test performance because, given the complexity of both the driving task and the training pattern, students will benefit from having more time to digest driving knowledge and skills via self-reflection and the absorption of task details. As such, we predict that a longer temporal gap between the end of training and the simulation test will improve the processing of the FI cues by students and will thus improve performance on the official driving skills test. While it is plausible that later simulation test dates would result in students forgetting essential details of their training experience, which could negatively impact official test performance, in our research context, the driving school generally ensures a short temporal gap between the end of training and the official test. Therefore, we believe that memory loss is unlikely to be a factor associated with simulation test scheduling in our context and hypothesize:

H3: The later the timing of the simulation test, the better the official test performance.

We further predict an interaction effect between feedback timing and the number of corrective FI cues on official driving skills test performance. In our research context, the real-time analysis of IoT data streams enables immediate feedback in the form of corrective FI cues that we believe influence student performance on the official test. This is one of the key features of smart services. Hypothesis 1 proposes that, through the learning mechanism, providing FI cues will have a positive impact on task performance. We therefore examine the impact of the number of corrective FI cues (i.e., the number of errors made in the simulation tests) on task performance. We propose that the positive impact of delayed feedback timing on official test performance will be stronger for students who make more errors during the simulation test. That is, the greater the number of errors made during the simulation test, the stronger the positive impact of delayed feedback timing on the official test performance. Following FIT, the number of corrective cues will correspond to the number of task details that students will reflect upon following the simulation test, and this process of reflection will be further integrated into their driving-skills knowledge base. Providing a high number of corrective FI cues directly after a student completes training may be overwhelming because of the lack of time available for digesting the training materials via self-reflection. In contrast, if the same number of corrective cues are provided later, after the student has had time to process the basic training materials, we predict that the student will be better able to use feedback received during the simulation test to learn from their errors, which will likely improve performance on the official test.

H4: The number of corrective FI cues provided by the simulation test will positively moderate the impact of feedback timing on official test performance.

3 Methodologies and Research Design

3.1 Methodology and Research Setting

We used a quasi-experimental method to test our hypotheses. The analysis unit of the study is the individual student of a driving school in an economically developed province in China. At present, this driving school has 30 different campuses, 44 training sites, 700 coaches, and about 40,000 students. In China, individuals need to pass four subject tests in order to obtain a driver's license. The Subject 1 (K1) and Subject 4 (K4) tests concern knowledge of driving laws and regulations. The Subject 2 (K2) test is a field test of driving skills and the Subject 3 (K3) test is a road test of driving skills. The tests are administered in order—for example, applicants must pass the Subject 1 test before they are allowed to take the Subject 2 test, and so on.

The driving school started using a K2 simulation testing system based on internet of vehicles (IoV) technology in July 2017. After completing the regular driving skills training program, students can make an appointment to use the simulation testing system. The simulation testing system fully simulates the experience and grading standards of the official examination. It can capture the driving action details of students in real time through onboard sensors and cameras and transfers information about students' driving errors to the driving school. After completing the simulation test, students immediately receive a feedback report (see sample feedback report in Appendix A). The simulation testing system is designed to replicate the actual format of the official K2 driving test so that students know what to expect in advance and so that student driving skills can be better assessed. Because not every student takes the simulation test after completing training, the actual effects of the simulation testing system can be accurately assessed.

3.2 Sample and Data

Since our data end in August 2017 and the school began using the simulation testing system in early July 2017, our sample period extends from May 2017 to August 2017, thus providing a temporally balanced data set (covering about two months before the introduction of the system and two months after). Our sample comprises all students enrolled in the K2 training program from May 2017 to August 2017. Regarding our quasi-experimental design, the sampled students differ along two dimensions: (1) when they took the official test (i.e., before versus after the introduction of the smart simulation testing system), and (2) the application of the "treatment" (i.e., taking versus not taking the simulation test). Our data include three parts: simulation test data, student information data, and training data summarized below (details are given in Appendix B):

1. Simulation test data are derived from the feedback report. They include the test date, test number, test score, error items, number of errors, deduction for each error, and so on.
2. Student information data include the age and gender of the student, where the training took place (i.e., on which campus of the school), and the passing date for each subject test.
3. Training data: After passing the K1 test, students train for the K2 test until they meet the training requirements of the driving school. Because students set their own schedules, each student's training schedule is slightly different in terms of total length of the training period (in days), length of each training session, interval between training sessions, etc. The training data include details about the training schedule of each student.

3.3 Variables

Our first research question investigates whether the simulation test affects student performance on the official driving skills test. Thus, our dependent variable is the test performance of each student. Technical reasons prevented us from accessing test scores for each student, but we do know when each student actually passed the official K2 driving skills test. Official test date, length of each student's training program, and the date of each student's simulation test are reported in Appendix B and were largely determined by the following scheduling procedures: After enrolling in driving school, students must register for an official K2 test date. The school then schedules the student's training sessions based on this date, making sure that the end of the training period is as close as possible to the official test date. Therefore, students who take the simulation test will do so shortly before they take the official test—i.e, sometime between the last training session and the official K2 test date. The basic rationale of training scheduling training is based on giving students the opportunity to take the official test while they still have a "fresh" memory of their training.

For our timing variable, we used the interval in days between the date of the last training session and the pass date of the official K2 driving skills test (defined as "Testday") as the dependent variable measuring the performance of students on the K2 test. We assume that longer intervals indicate lower performance on the K2 test because longer intervals suggest that the student may have taken the test multiple times before passing it. To test whether feedback timing affected test performance for students taking the simulation test, we examined how long after taking the simulation test it took a student to pass the K2 test. So, the dependent variable we used is "Ptestday," which equals the interval between the pass date of the K2 test and the simulation test date divided by the interval between the pass date of the K2 test and the date of the last training session.

Our independent variables are: (1) whether the student took the simulation test, and (2) feedback timing. Whether the student took a simulation test is defined as "Simulation"—a dummy variable equals 1 if the student took the simulation test and 0 if the student did not take it. We used "Feedbackday" to measure the feedback timing of the simulation test. "Feedbackday" is the interval in days between the simulation test date and the date of the last training session. If the student took multiple simulation tests, we treated the median date of all the simulation test dates as the student's simulation test date.

Beyond this, we assessed whether the training schedule affected the relationship between the simulation test and official test performance. We used "Gapstd," the standard deviation of the interval between different training sessions (in minutes), and "Gapmean," the mean of the interval between different training sessions (in minutes), to measure the different training schedules. We also tested whether feedback information affected the relationship between feedback timing and test performance. We used "Falsenum_mean" to measure the feedback information of the simulation test. "Falsenum_mean" is the mean number of errors made in each simulation test taken by a student (some students took multiple simulation tests).

Many factors determine a student's likelihood of taking one or more simulation tests and these factors may also correlate with official test performance. Our estimated effect of the simulation test on official test performance is subject to selection biases. To mitigate this concern, we matched each treated student (students who took one or more simulation tests) with a control student who did not take the simulation test and used the matched sample throughout our regression analyses. We used the propensity score matching (PSM) method to construct our matched sample. We estimated the following logit model using all student data following the use of the simulation testing system (July 2017 and August 2017):

$$Logit(Simulation_i) = \alpha + \beta_1 * Ln(Age_i) + \beta_2 * Female_i + \beta_3 * Ln(Trainday_i) + \beta_4 * Ln(Sub2period_i) + control(School\ campus, Time) + \varepsilon \quad (1)$$

where $Simulation_i$ equals 1 if student i took the simulation test and 0 otherwise. We controlled for the log of student age ($Ln(Age_i)$), gender ($Female_i$), the log of total training days ($Ln(Trainday_i)$), the log of total duration of each training session ($Ln(Sub2period_i)$), and *School campus* and *Time* as the factors that might affect a student's likelihood of taking a simulation test. See Appendix C for definitions of the variables.

3.4 Summary Statistics

Table 1 presents the summary statistics of our final sample consisting of 2,812 students. All continuous variables were winsorized at the 1% level at both tails of their distributions. Beyond this, we took the logarithm of all the continuous variables with absolute values to mitigate the influence of distribution skewness and centralized them to mitigate the effect of multicollinearity between the variables after adding the interaction terms. Panel A of Table 1 presents the summary statistics of the variables after winsorization. Panel B of Table 1 presents the summary statistics of the continuous variables after centering.

In our final sample, 62.2% of the students were female. The average age of the students was 29. It took, on average, 38 days after the last training session to pass the K2 test. On average, students trained for 1187 minutes (19.78 hours) over 25 training days and the mean interval between each training session was 9492 minutes (6.59 days). Among students who took the simulation test, the mean interval between the simulation test date and the date of the last training session was 34 days, and during each simulation test, students made an average of 3.57 errors. The correlation coefficient matrix of each variable in our regression models is shown in Table 2.

4 Empirical Results

4.1 The Baseline Model

Our baseline regression specification is written as follows:

$$Ln(Testday_i) = \alpha + \beta_1 * Simulation_i + Controls + \varepsilon \quad (2)$$

where $Ln(Testday_i)$ is the log of the time interval between the last training session and the K2 test pass date in days. $Simulation_i$ is a dummy variable that equals 1 if student i took simulation test and 0 otherwise. We controlled the log of student age ($Ln(Age_i)$), gender ($Female_i$), log of total training days ($Ln(Trainday_i)$), and log of the total duration of each training session ($Ln(Sub2period_i)$) in our regression. We also included campus-fixed effects to control for the impact of unobservable campus characteristics; time-fixed effects are included to account for the aggregate time variation in K2 test performance.

We present baseline regression results in Table 3, where Column 1 presents the results without control variables and Column 2 presents the results with control variables. Both columns include campus- and time-fixed effects. The coefficients of *Simulation* in both columns are negative and statistically significant (*t*-statistics = -2.849 and -2.883 in Columns 1 and 2, respectively), suggesting that, on average, compared with students who did not take simulation tests, students who took the simulation test passed the K2 test 1.09 days (=$e^{0.09}$) sooner. The coefficients of the control variables show that, on average, the older the student, the more time it took to pass the test. As a group, females passed the test more quickly than males, and, on average, the more days a student trained, the more quickly he or she passed the test. The incremental percentage of adding the *Simulation* variables of R^2 is around 1.89%—in other words, adding the *Simulation* variables increased the R^2 of the regression model from 13.21% to 13.46%.

In order to further address the potential endogeneity issue, we also selected the shortest distance between each training school campus and the main campus as the instrumental variable (IV) and conducted two-stage least squares (2SLS) analysis. The simulation system was installed at the main campus only; trainees could decide whether or not to take simulation test based, in part, on the distance between their respective training school campus and the main campus.

This distance, however, would not affect trainee test performance, so the distance is related to simulation possibility and is orthogonal to the test performance. Before conducting the analysis, we first used the sample following PSM and conducted the test of endogeneity (DWH test). Results showed that the variables are exogenous, suggesting that the endogenous problem in our sample is not serious.

Table 1. Summary Statistics

Variables	***N***	**Mean**	***SD***	**Min**	**Median**	**Max**
Panel A summary statistics of the variables after winsorization						
Testday	2812	37.750	33.340	3.000	25.000	160.000
Simulation	2812	0.500	0.500	0.000	0.500	1.000
Gapmean	2812	9492.000	11000.000	802.000	5502.000	62000.000
Gapstd	2812	9956.000	15000.000	7.778	4400.000	90000.000
Age	2812	29.050	7.971	18.000	28.000	50.000
Female	2812	0.622	0.485	0.000	1.000	1.000
Trainday	2812	25.020	27.220	3.000	15.000	152.000
Sub2period	2812	1187.000	446.200	693.000	1073.000	5060.000
Ptestday	1406	0.100	0.120	0.006	0.063	0.667
Feedbackday	1406	34.200	33.100	1.000	22.000	173.000
Falsenum_mean	1406	3.573	1.742	1.000	3.250	9.500
Panel B summary statistics of the variables after logarithm and centering						
Ln(Testday)	2812	0.000	0.855	-2.178	-0.058	1.798
Ln(Gapmean)	2812	0.000	1.021	-1.954	-0.028	2.400
Ln(Gapstd)	2812	0.000	1.612	-6.210	0.128	3.141
Ln(Age)	2812	0.000	0.269	-0.442	0.000	0.580
Ln(Trainday)	2812	0.000	0.943	-1.669	-0.059	2.256
Ln(Sub2period)	2812	0.000	0.310	-0.485	-0.048	1.503
Ptestday	1406	0.000	0.120	-0.095	-0.038	0.566
Ln(Feedbackday)	1406	0.000	0.946	-3.118	-0.027	2.036
Ln(Falsenum_mean)	1406	0.000	0.497	-1.156	0.023	1.096

Notes: The sample includes both students who took the simulation test after July 2017 and the matched students who did not take the simulation test. We winsorized all the continuous variables at a 1% level at both tails. Panel A presents the summary statistics of the variables after winsorization. We then took the logarithm of all the continuous variables with absolute values to mitigate the influence of distribution skewness, and centralized all variables used in our regressions to mitigate the effect of multicollinearity between the variables after adding the interaction terms. Panel B presents the summary statistics of the continuous variables after centering

Table 2. Correlation Table

	Ln(Testday)	Simulation	Ln(Gapmean)	Ln(Gapstd)	Ln(Age)	Female	Ln(Trainday)	Ln(Sub2period)
Ln(Testday)	1							
Simulation	-0.045**	1						
Ln(Gapmean)	-0.035*	0.029	1					
Ln(Gapstd)	-0.067***	0.018	0.838***	1				
Ln(Age)	0.180***	-0.024	0.065***	0.052***	1			
Female	-0.147***	-0.030	-0.028	-0.001	0.038**	1		
Ln(Trainday)	-0.075***	0.003	0.897***	0.840***	0.079***	-0.011	1	
Ln(Sub2period)	-0.005	0.043**	-0.051***	0.058***	0.030	-0.039**	0.136***	1

Notes: This table presents the correlation coefficient matrix of each variables in our regression models. The symbols ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

Table 3. Effect of the Simulation Test on K2 Test Performance

	1 **Ln(Testday)**	2 **Ln(Testday)**
Simulation	-0.093*** (-2.849)	-0.091*** (-2.883)
Ln(Age)		0.605*** (10.541)
Female		-0.278*** (-8.792)
Ln(Trainday)		-0.056*** (-3.369)
Ln(Sub2period)		-0.085 (-1.475)
Intercept	-0.043 (-0.720)	0.099 (1.571)
Campus- and time-fixed effects	Yes	Yes
No. of Observations	2812	2812
***Adj. R*-square**	0.075	0.135

Notes: The sample includes both students who took the simulation test after July 2017 and the matched students who did not take the simulation test. *Ln(Testday)* is the log of time interval between the last training session and the K2 test pass date in days. *Simulation* is a dummy variable that equals 1 if student *i* took the simulation test and 0 otherwise. *Ln(Age)* is the log of student age, *Female* is a dummy variable that equals 1 if the student is female. *Ln(Trainday)* is the log of total days between the date of the first training session and the date of the last training session. *Ln(Sub2period)* is the log of the sum of each training period in minutes. We included campus-fixed effects to control for the impact of unobservable campus characteristics and time-fixed effects to account for the aggregate time variation in K2 test performance. The *t*-statistics are in parentheses. The symbols ***, **, * denote significance at the 1%, 5%, and 10% levels, respectively.

Following this, we conducted the DWH test using the sample prior to PSM; the results are presented in Table 4. For the first-stage result, the longer the distance, the less likely it was that students would take the simulation test. For the second-stage result, the coefficient of $\widehat{Simulation}$ remains significantly negative, suggesting that students who took the simulation test generally passed the K2 test more quickly.

4.2 Difference-in-Difference Analysis

To alleviate endogeneity concerns about our baseline results, we used difference-in-difference (DID) analysis to test whether the simulation testing system improved K2 test performance. The external shock we employed was the use of the simulation testing system beginning in July 2017. Therefore, we defined the two months prior to the introduction of the simulation testing system (May-June 2017) as the "before" period and the two months after the introduction of the simulation testing system (July-August 2017) as the "after" period.

Table 4. Effect of the Simulation Test on K2 Test Performance: IV Test

Dependent variable	1 first-stage simulation	2 second-stage Ln(Examday)
Distance	-0.003*** (-16.136)	
$\widehat{Simulation}$		-0.659*** (-3.678)
Ln(Age)	0.015 (1.465)	0.464*** (16.704)
Female	-0.003 (-0.655)	-0.262*** (-17.556)
Ln(Trainday)	-0.002 (-0.709)	-0.133*** (-16.713)
Ln(Sub2period)	0.048*** (5.566)	0.065** (2.423)
Intercept	0.002 (0.557)	-0.031* (-1.720)
Time fixed effects	Yes	Yes
No. of Observations	12850	12850
Adj. R-square	0.160	0.030

Notes: The sample includes all the students who pass the K2 exam between May 2017 and August 2017. *Ln(Testday)* is the log of time interval between the last training session and the K2 test pass date in days. *Simulation* is a dummy variable that equals 1 if student *i* took the simulation test and 0 otherwise. *Distance* the instrumental variable (IV), which is defined as the shortest distance between students' respective training school campuses and the main campus. *Ln(Age)* is the log of student age, *Female* is a dummy variable that equals 1 if the student is female. *Ln(Trainday)* is the log of total days between the date of the first training session and the date of the last training session. *Ln(Sub2period)* is the log of the sum of each training period in minutes. We included time-fixed effects to account for the aggregate time variation in K2 test performance. The t-statistics are in parentheses. The symbols ***, **, * denote significance at the 1%, 5%, and 10% levels, respectively.

During the before period, no individuals took the simulation test because there was no simulation testing system available. So, we defined the "treatment" and "control" samples based on the campus on which students were enrolled. If students had access to the simulation test in July 2017 and later, then we treated all the students as the "treatment" group. Students that did not take the simulation test in July 2017 or later were treated as the "control" group. Also, to mitigate the selection bias concern, we used the PSM method to construct our sample. The logistic model to match the control sample with the treatment sample is as follows:

$$Logit(Campussimu_i) = \alpha + \beta_1 * Ln(Age_i) + \beta_2 * Female_i + \beta_3 * Ln(Trainday_i) + \beta_4 * Ln(Sub2period_i) + control(Time) + \varepsilon \quad (3)$$

where *Campussimu*$_i$ is a dummy variable that equals 1 if students on a specific campus took the simulation test after July 2017 and 0 otherwise. The control variables in Model (3) are the same as Model (2). Because *Campussimu*$_i$ is highly correlated with the school campus dummy, we only included time-fixed effects in this model.

The DID analysis results presented in Table 5 show that, on average, it took more time for both the treatment and the control groups to pass the K2 test after implementation of the simulation testing system. However, in the two months following the introduction of the simulation test, the difference in *Testday* between the treatment group and the control group changed from 1.395 days to -0.526 days, suggesting that, as a whole, students taking the simulation test passed the test more quickly than students who did not. The difference is significant at a 10% level (single-tail test). The analysis confirms our baseline results that the simulation testing system improves test performance.

Table 5. Difference-in-Difference Analysis

	Testday		
	Before	**After**	**Diff-in-diff**
Control	29.577	37.887	
Treated	30.972	37.361	
Diff (T-C)	1.395	-0.526	-1.921*

Notes: The sample includes matched treatment and control groups using the PSM method. The original treatment group comprised students on campuses that had access to the simulation test in July 2017 and later. The original control group was made up of students on campuses that did not have access to the simulation test in July 2017 or later. We matched the original control group with the treatment group using the PSM method. The *Before* period includes the two months before the simulation testing system was implemented, May 2017 and June 2017. The *After* period includes the two months after the simulation testing system was implemented, July 2017 and August 2017. The symbols ***, **, * denote significance at the 1%, 5%, and 10% levels, respectively. Our concern is whether the diff-in-diff is significantly negative, so the difference test is a one-tail test.

4.3 Moderating Effects of Training Schedule Differences

To investigate whether different student training schedules affected the relationship between the simulation test and test performance, we estimated the following model:

$$Ln(Testday_i) = \alpha + \beta_1 * Simulation_i + \beta_2 * Ln(Gapmean_i)\ \big(or\ Ln(Gapstd_i)\big) + \beta_3 * Simulation_i \times Ln(Gapmean_i)\ \big(or\ Simulation_i \times Ln(Gapstd_i)\big) + Controls + \varepsilon \qquad (4)$$

where $Ln(Gapmean_i)$ is the log of mean of interval in minutes between each training session and $Ln(Gapstd_i)$ is the log of standard deviation of interval in minutes between each training session. We determined that a smaller mean of the gaps between training sessions indicated more intensive training, and that a smaller standard deviation of the gaps between training sessions indicated more regular training. We used these two variables to measure student training schedules. Control variables are the same as the variables in Model (2). We also included campus- and time-fixed effects in our regressions.

Table 6 presents the regression results of Model (4). The variable used to measure different training schedules in Columns 1 and 2 is $Ln(Gapmean_i)$ and $Ln(Gapstd_i)$ in Columns 3 and 4. Columns 1 and 3 are results without control variables and Columns 2 and 4 present the results with control variables. All columns include campus- and time-fixed effects. The coefficients of interaction terms are all negative but insignificant, suggesting that, on average, different training schedules had no significant effect on the relationship between the simulation test and official test performance. The R^2 of the regression models increased by 0.52% and 0.15%, respectively, after adding the moderating variables $Ln(Gapmean_i)$ and $Ln(Gapstd_i)$.

4.4 Feedback Timing and Test Performance

To investigate whether feedback timing affects test performance for the students who took the simulation test, we ran the following regression model:

$$Ptestday_i = \alpha + \beta_1 * Ln(Feedbackday_i) + Controls + \varepsilon \qquad (5)$$

Where $Ptestday_i$ measures the time it takes to pass the K2 test after the simulation test and equals the interval between the K2 test pass date and the simulation test date divided by the interval between the K2 test pass date and the date of the last training session. $Ln(Feedbackday_i)$ is the log of the interval in days between the simulation test date and the date of the last training session. Control variables are the same as the variables in Model (2), and we also included campus- and time-fixed effects in our regression. Our sample here comprises the students who took the simulation tests in our final sample.

Table 7 presents the regression results of Model (5). Column 1 presents the results without control variables and Column 2 presents the results with control variables. All columns include campus- and time-fixed effects. The coefficients of *Ln(Feedbackday)* are all significantly negative, suggesting that, on average, the longer the interval between the simulation test and the last training session, the less time it took to pass the official driving skills test after taking the simulation test. It takes time for students to digest skills learned during the training period. Therefore, we suspect that the long interval between the simulation test and final training session gave students enough time to fully understand and digest the driving skills they learned, allowing them to then quickly pass the official test. Compared with the models using control variables only, adding *Ln(Feedbackday)* increased the R^2 of the regression model by around 217.17%.

Table 6. Moderating Effect of Different Training Schedules

	1 Ln(Testday)	2 Ln(Testday)	3 Ln(Testday)	4 Ln(Testday)
Simulation	-0.092*** (-2.822)	-0.093*** (-2.917)	-0.092*** (-2.802)	-0.091*** (-2.871)
Ln(Gapmean)	-0.021 (-0.915)	0.065 (1.469)		
Simulation×Ln(Gapmean)	-0.013 (-0.436)	-0.015 (-0.508)		
Ln(Gapstd)			-0.025* (-1.748)	-0.006 (-0.320)
Simulation×Ln(Gapstd)			-0.003 (-0.135)	-0.008 (-0.454)
Ln(Age)		0.607*** (10.574)		0.604*** (10.521)
Female		-0.278*** (-8.781)		-0.278*** (-8.796)
Ln(Trainday)		-0.114** (-2.416)		-0.040 (-1.347)
Ln(Sub2period)		-0.052 (-0.838)		-0.089 (-1.533)
Intercept	-0.037 (-0.623)	0.090 (1.432)	-0.041 (-0.687)	0.098 (1.567)
Campus- and time-fixed effects	Yes	Yes	Yes	Yes
No. of Observations	2812	2812	2812	2812
Adj. R-square	0.077	0.135	0.078	0.135

Notes: The sample includes both students who took the simulation test after July 2017 and matched students who did not take the simulation test. *Ln(Testday)* is the log of the time interval between the last training session and the K2 test pass date in days. *Simulation* is a dummy variable that equals 1 if student *i* took the simulation test and 0 otherwise. *Ln(Gapmean)* is the log of the mean interval in minutes between each training session. *Ln(Gapstd)* is the log of the standard deviation in minutes of the interval between each training session. *Ln(Age)* is the log of student age, *Female* is a dummy variable that equals 1 if the student is a female. *Ln(Trainday)* is the log of total days between the first training session and the last training session. *Ln(Sub2period)* is the log of the duration of each training session in minutes. The variable used to measure different training schedules in Columns 1 and 2 is *Ln(Gapmean)* and *Ln(Gapstd)* in Columns 3 and 4. Columns 1 and 3 are results without control variables and Columns 2 and 4 present the results with control variables. All columns include campus- and time-fixed effects. The t-statistics are in parentheses. The symbols ***, **, * denote significance at the 1%, 5%, and 10% levels, respectively.

Table 7. The Effect of Feedback Timing on Test Performance

	Column 1 Ptestday	Column 2 Ptestday
Ln(Feedbackday)	-0.080*** (-20.951)	-0.081*** (-20.886)
Ln(Age)		0.014 (1.536)
Female		-0.014*** (-2.839)
Ln(Trainday)		0.009*** (3.416)
Ln(Sub2period)		-0.002 (-0.298)
Intercept	0.051*** (5.131)	0.059*** (5.453)
Campus- and time-fixed effects	Yes	Yes
No. of Observations.	1406	1406
Adj. R-square	0.491	0.500

Notes: The sample includes students who took the simulation tests in our final sample. *Ptestday* is the interval between the K2 test pass date and the simulation test date divided by the interval between the K2 test pass date and the date of the last training session. *Ln(Feedbackday)* is the log of the interval in days between the simulation test date and the date of the last training session. *Ln(Age)* is the log of student age, *Female* is a dummy variable that equals 1 if the student is a female. *Ln(Trainday)* is the log of total days between the first training session and the last training session. *Ln(Sub2period)* is the log of the duration of each training session in minutes. Column 1 presents results without control variables and Column 2 us the results with control variables. All columns include campus- and time-fixed effects. The t-statistics are in parentheses. The symbols ***, **, * denote significance at the 1%, 5%, and 10% levels, respectively.

Table 8. Moderating Effect of Feedback Information

	1 Ptestday	2 Ptestday	3 Ptestday	4 Ptestday
Ln(Feedbackday)	-0.080*** (-21.156)	-0.081*** (-21.054)	-0.071*** (-13.418)	-0.072*** (-13.676)
Ln(Falsenum_mean)	0.013*** (2.781)	0.010** (1.978)		
Ln(Feedbackday) × Ln(Falsenum_mean)	-0.020*** (-2.794)	-0.020*** (-2.861)		
Falsenumdummy			0.011** (2.465)	0.008* (1.772)
Ln(Feedbackday) × Falsenumdummy			-0.017** (-2.242)	-0.016** (-2.221)
Ln(Age)		0.012 (1.376)		0.013 (1.444)
Female		-0.012** (-2.428)		-0.012** (-2.497)
Ln(Trainday)		0.009*** (3.352)		0.009*** (3.286)
Ln(Sub2period)		-0.002 (-0.225)		-0.002 (-0.215)
Intercept	0.044*** (4.579)	0.053*** (4.872)	0.040*** (4.076)	0.050*** (4.489)
Campus- and time-fixed effects	Yes	Yes	Yes	Yes
No. of Observations	1406	1406	1406	1406
Adj. R-square	0.500	0.508	0.497	0.505

Notes: The sample includes students who took the simulation tests in our final sample. *Ptestday* is the interval between the K2 test pass date and the simulation test date divided by the interval between the K2 test pass date and the date of the last training session. *Ln(Feedbackday)* is the log of interval days between the simulation test date and the date of the last training session. *Ln(Falsenum_mean)* is the log of the mean number of errors made during the simulation tests. *Falsenumdummy* is a dummy variable that equals 1 if the mean of errors during the simulation tests (*Falsenum_mean*) of student *i* is larger than the median of the sample (treatment sample) and 0 otherwise. *Ln(Age)* is the log of student age, *Female* is a dummy variable that equals 1 if the student is a female. *Ln(Trainday)* is the log of total days between the date of the first training session and the date of the last training session. *Ln(Sub2period)* is the log of the sum of each training period in minutes. Columns 1 and 3 are results without control variables and Columns (2) and (4) present the results with control variables. All columns include campus- and time-fixed effects. The t-statistics are in parentheses. The symbols ***, **, * denote significance at the 1%, 5%, and 10% levels, respectively.

4.5 Moderating Effects of Feedback Information

To assess whether the feedback information affected the relationship between feedback timing and test performance, we used *Ln(Falsenum_mean)*, which is the log of the mean number of errors made during the simulation tests, in our regression model to measure the feedback information. Our regression model is as follows:

$$Ptestday_i = \alpha + \beta_1 * Ln(Feedbackday_i) + \beta_2 * Ln\left(Falsenum_{mean_i}\right) + \beta_1 * Ln(Feedbackday_i) \times Ln\left(Falsenum_{mean_i}\right) + Controls + \varepsilon \quad (6)$$

Table 8 presents the regression results of Model (6). Column 1 presents the results without control variables and Column 2 presents the results with control variables. All columns include campus- and time-fixed effects. The coefficients of *Ln(Feedbackday)* are both significantly negative and the coefficients of the interaction term *Ln(Feedbackday) × Ln(Falsenum_mean)* are both significantly negative. The results show that the more errors the student made in the simulation test, the stronger the negative relation between feedback timing and test performance. We replaced *Ln(Falsenum_mean)* with *Falsenumdummy* and ran Model (6) again. *Falsenumdummy* is a dummy variable that equals 1 if the mean of errors during the simulation tests (*Falsenum_mea*n) of student *i* is larger than the median of the sample (treatment sample), and 0 otherwise. The results are shown in Columns 3 and 4 of Table 8. Furthermore, we determined that the coefficients of *Ln(Feedbackday)* and the interaction terms are all significantly negative, confirming our results. Compared with the model with control variables and main effect, adding moderating variables increased the R^2 of the regression model by around 1.58% for *Ln(Falsenum_mean)* and 1.00% for *Falsenumdummy*. Presumably, students who make higher numbers of errors in the simulation test have relatively poor driving skills. Their need to spend time digesting and absorbing the information learned during training may therefore be higher. We thus suspect that the digestion effect of feedback time is more significant for these students.

5 Discussion

5.1 Theoretical Implications

Drawing on the theory of feedback intervention, the current research examines the impact of a smart simulation testing system on individual students in the context of a driving school supported by internet of vehicles (IoV) technology. Our research makes a number of theoretical contributions to the smart services literature, in general, and to research on the impact of smart technologies on individual users' task performance, in particular.

First, we contribute to the literature on service science by focusing on the *smart* element enabled by the internet of things (IoT) used in a service context. We postulate that the data streams generated in IoT technologies can serve as the foundation to invent new and smart services that at least partially replace human intelligence in a standard service setting and, more importantly, exert significant influence on performance outcomes. In our research context, the simulation test is a standard service setting where driving students perform a predefined set of actions during the course of a standard test and receive standard evaluations of their task performance. The service is *smart* because it provides feedback that is usually offered by human coaches based on real-time data streams from sensors and cameras installed in a vehicle using IoV technology. Real-time data analytics provide feedback on both overall task performance and diagnostic information about task details. Our empirical study demonstrates that feedback from the smart service significantly impacts student performance on the official driving skills field test necessary to receive a driver's license in China. Our research thus provides initial evidence of the value of smart services in this context.

Second, we focus on the impacts of smart IoT technologies on individual users and introduce feedback intervention theory (FIT) as the overarching theoretical foundation for examining the impacts of IoT-based smart services on individuals' task performance. Following FIT, we theorize two mechanisms underlying the hypothesized effect: the motivation mechanism and the learning mechanism, which correspond, respectively, to the two types of FI cues—velocity cues and corrective cues. In our research context, these two types of cues take the form of overall simulation test scores and the specific errors made during the test, respectively. We theorize that this feedback from the smart test-simulation system has a positive impact on students' official driving skills test performance. Our results support the validity of these hypothesized impacts.

Third, we adapt FIT and propose feedback timing as an antecedent of task performance. Moreover, we also propose an interaction effect between feedback timing and the number of feedback cues related to task details. Feedback timing and the number of feedback cues were not explicitly discussed in the original FIT framework. However, in the context of a smart service, the rich data streams generated from IoT technology can enable a large quantity of feedback cues at a granular level. Also, the real-time nature of data streams implies that feedback can be provided anytime. Therefore, both the timing and the amount of feedback should be optimized to fit individuals' ability to process the information. Following the fundamental notion of task-details learning in FIT, we developed hypotheses regarding the impact of feedback timing and its interaction with the number of feedback cues in the context of the smart simulation testing system. Feedback timing is defined as the time elapsed between the end of driving school training and the simulation test event, while the number of feedback cues comprises the number of errors made by a student during the simulation test. We found that both feedback timing and its interaction with the number of cues significantly influenced student performance on the official driving skills test.

5.2 Managerial Implications

Our findings also have managerial implications for the design and management of IoT-based smart services—for example, the use of real-time data streams from IoV for the purpose of smart service innovation.

First, our research shows that smart services can offset the high costs of human capital by partially replacing human intelligence—in our context, through replacing human coaching during the simulation test. According to our interviews with both the director of the driving school and coaching representatives, simulation testing offers an effective means of better preparing students for the official driving skills test. The smart system provides students with additional opportunities to perform tasks in an authentic testing scenario, which effectively facilitates the self-diagnosis of their driving skills. Therefore, this system improves the official test performance of driving school students, increases the turnover volume of the driving school, and saves human time and energy, which can then be invested in the training of additional students. Finally, according to our interviews with the students, they perceived the simulation testing system to be a sufficiently *smart* and important complement to traditional face-to-face coaching methods because it offers students additional learning options and thereby contributes to a diverse learning environment.

Second, our results also have implications for designing smart services, in general, and smart feedback, in particular. Our research demonstrates not only the importance of the *form* of feedback (overall score vs. detailed summary of errors) for influencing performance outcomes (official test success), but show that the *timing* and the *quantity* of feedback are also crucial for performance outcomes. Therefore, developers and managers of smart IoT applications should pay attention to all these key design factors to make their services/feedback *smarter*—i.e., more personalized toward the heterogeneous needs of users. The key is to reach an optimal trade-off between giving users enough time to adequately digest and process their training while also making sure that user memory is fresh and active enough to effectively perform critical tasks.

6 Limitation and Future Research Directions

Our research has a number of limitations. First, this research primarily focuses on the context of IoV-based smart services (simulation tests). Future research is needed to evaluate the generalizability of our research in other IoT scenarios—for example, in the area of health care, our research could be used to test the impact of feedback interventions on patient attempts to integrate healthy habits. Second, our research model only covers part of the FIT framework. Future research could examine the role of other FIT elements such as higher-level motivations (i.e., self-realization—the self level in feedback intervention theory) in the context of IoT-based smart services. Third, we adopted a quasi-experimental design in our study based on the nature of the secondhand data we used. We adopted the PSM method to minimize possible causality issues such as sample heterogeneity; however, we cannot rule out other alternative explanations—for example, unobservable traits like motives and ability. Students who take the official test more seriously, for instance, might want to be better prepared and thus might engage in off-the-record self-training or might sign up for the simulation test even though they are training at a campus that does not offer simulation testing.[3] In addition, while the timing of the simulation test (feedback timing) is generally scheduled to be as close as possible to the official test date, there still exists the potential endogeneity issue resulting from eager driving students or those with hard time constraints seeking an earlier simulation test date. In such cases, then the empirical analysis of the interaction effect between feedback timing and other variables might not be causal. Future research could employ a random-field-experiment approach to better asses causality inferences. Finally, while using official test scores for all the students as the performance metric of the dependent variable would certainly be desirable, government regulations prevented us from accessing these data, which therefore necessitated the current temporal proxy of official test performance. We call for future research with richer field data to provide a more direct measure of official test performance and thus a more direct test of our hypotheses.

In the specific context of driver training, a more traditional method of offering feedback is in a face-to-face (FTF) setting—i.e., having a coach sit next to a student during a simulation test to provide immediate feedback. To assess whether FTF feedback is used during the simulation testing at the focal driving school, we interviewed the director of the driving school and learned that students on all campuses receive FTF feedback from their coach during a simulated driving test. The IoV-based smart system we examine here thus offers additional feedback following completion of the training program and before taking the official test. As such, we can still consider the IoV-enabled simulation test as a "treatment," with the comparison we make here being essentially one between FTF and FTF+IoV. Since the research setting and the quasi-experimental design of the current research do not allow us to make a direct comparison between FTF and IoV, we call on future research with randomized field experiments to make this direct comparison.

Acknowledgments

We thank Kenneth Hsing Cheng, Ting-Peng Liang, Chih-Hung Peng, Suprateek Sarker, Yu-chen Yang, the anonymous reviewers, and the participants of the 2018 PolyU-AIS Research Workshop, 2018 Advanced Electronic Commerce Forum: Big Data for Smart Commerce and Service Innovation for helpful discussions and useful suggestions. Xin Xu acknowledges the financial support of the PolyU iDEAS Grant (Project No. 1-99XZ). Yong Jin also acknowledges the financial support of the PolyU Central Research Grant (Project No. G-YBZV). Wen Xie and Xin Xu are co-first authors. Ruiqi Liu and Yong Jin are corresponding authors. All errors are our own.

[3] We thank both of the reviewers for this valuable point.

References

Butler, A. C., Karpicke, J. D., & Roediger, H. L. (2007). The effect of type and timing of feedback on learning from multiple-choice tests. *Journal of Experimental Psychology*, *13*(4), 273-281.

Fleishman, E. A. (1972). On the relation between abilities, learning, and human performance. *American Psychologist*, *27*(11), 1017-1032.

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, *119*(2), 254-284.

Lee, I., & Lee, K. (2015). The internet of things (IoT): Applications, investments, and challenges for enterprises. *Business Horizons*, *58*(4), 431-440.

Müller L., Rivera-Pelayo V., & Heuer S. (2012). Persuasion and reflective learning: Closing the feedback loop. In M. Bang & E. L. Ragnemalm (Eds.), *Persuasive technology: Design for health and safety*. Springer.

Porter, M. E., & Heppelmann, J. E. (2015). How smart, connected products are transforming companies. *Harvard Business Review*, *93*(10), 96-16.

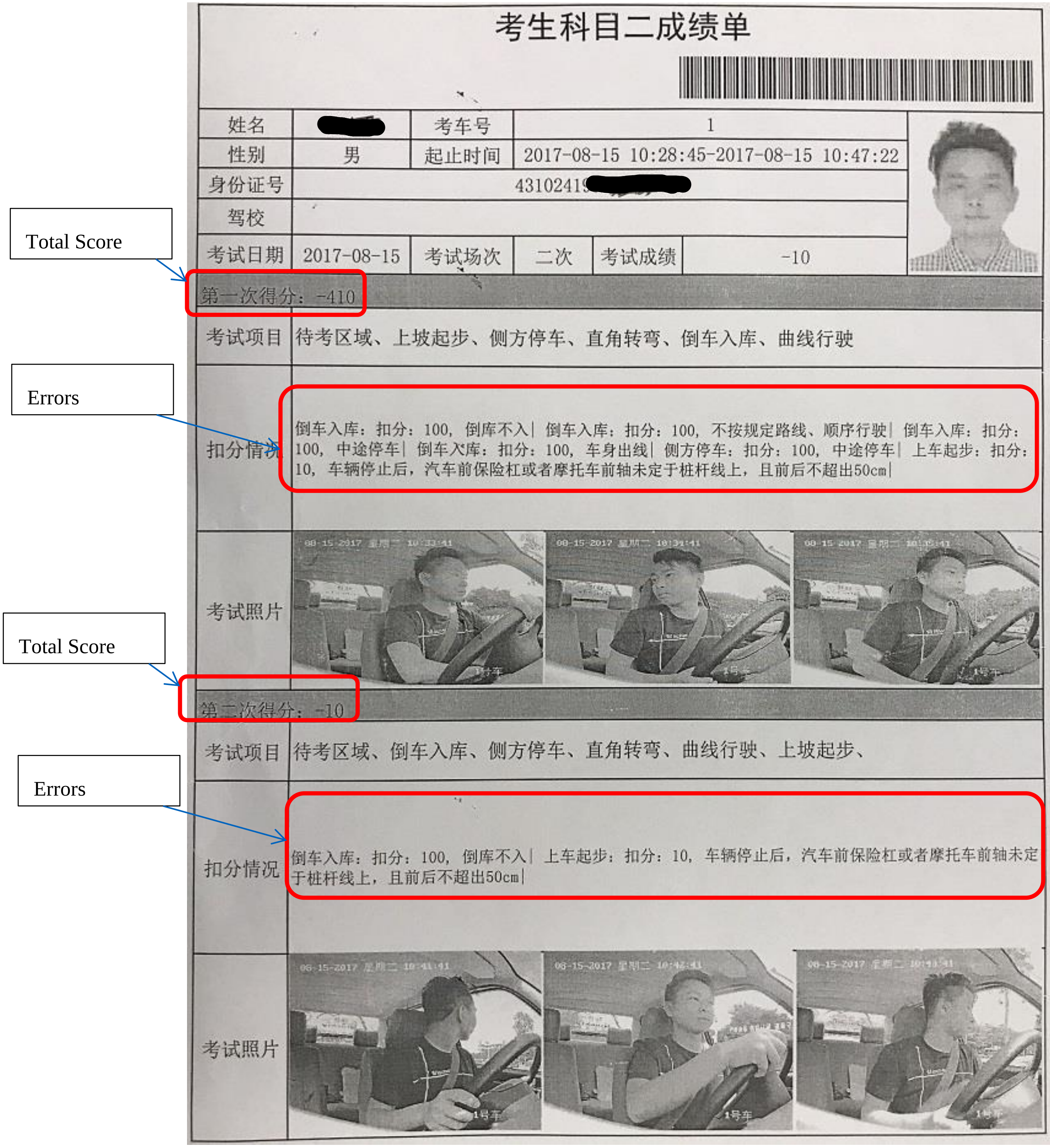
Rivera-Pelayo, V., Zacharias, V., Müller, L., & Braun, S. (2012). Applying quantified self approaches to support reflective learning. *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*.

Soleymanian, M., Weinberg, C., & Zhu, T. (2016). *The value of usage-based insurance beyond better targeting: better driving* (Working paper, Sauder School of Business, University of British Columbia). https://pdfs.semanticscholar.org/a750/a06e7e66783535f160d2ae69240506dcccce.pdf?_ga=2.259607689.266783334.1592350469-909906619.1592350469.

Thornock, T. A. (2016). How the timing of performance feedback impacts individual performance. *Accounting, Organizations and Society*, *55*, 1-11.

Wortmann, F., & Flüchter, K. (2015). Internet of things. *Business & Information Systems Engineering*, *57*(3), 221-224.

Appendix A



考生科目二成绩单

姓名	[redacted]	考车号	1		
性别	男	起止时间	2017-08-15 10:28:45-2017-08-15 10:47:22		
身份证号	4310241[redacted]				
驾校					
考试日期	2017-08-15	考试场次	二次	考试成绩	-10

第一次得分：-410

考试项目	待考区域、上坡起步、侧方停车、直角转弯、倒车入库、曲线行驶
扣分情况	倒车入库：扣分：100，倒库不入| 倒车入库：扣分：100，不按规定路线、顺序行驶| 倒车入库：扣分：100，中途停车| 倒车入库：扣分：100，车身出线| 侧方停车：扣分：100，中途停车| 上车起步：扣分：10，车辆停止后，汽车前保险杠或者摩托车前轴未定于桩杆线上，且前后不超出50cm|
考试照片	

第二次得分：-10

考试项目	待考区域、倒车入库、侧方停车、直角转弯、曲线行驶、上坡起步、
扣分情况	倒车入库：扣分：100，倒库不入| 上车起步：扣分：10，车辆停止后，汽车前保险杠或者摩托车前轴未定于桩杆线上，且前后不超出50cm|
考试照片	

Figure A1. Sample Driving-Simulation Report

Appendix B

Table B1. Data Structure

Name	Type	Description
Part A: Simulation test data		
Id	char	Encoding of the simulation test
Id_card_md5	char	Postencryption encoding of student's ID card number
Car_type	char	Vehicle type
Test_no	tinyint	Simulation test number
Score	smallint	Simulation test score
Test_date	date	Simulation test date
Test_id	char	Encoding of the simulation test
Test_item	varchar	Test item (total of 6 items in Subject 2 test)
Deduction	smallint	Deduction from score for each error
Reason	text	Deduction reason for each error
Part B: Student information data		
Id_card_md5	varchar	Postencryption encoding of student's ID card number
Subject1_date	date	Subject 1 test pass date
Subject2_date	date	Subject 2 test pass date
Long_way_date	date	Long-Way test pass date
Subject3_date	date	Subject 3 test pass date
Safe_training	date	Safe-Training test pass date
Student	varchar	Name of student
Gender	varchar	Gender of student
Birthday	date	Student date of birth
Age	varchar	Age of student
Training_campus	date	Campus where student was enrolled
Part C: Training data		
Id	int	ID number of training record
Student	varchar	Name of student
Id_card_md5	char	Postencryption encoding of student's ID card number
Gender	varchar	Gender of student
Birthday	date	Student date of birth
Age	varchar	Age of student
Training_campus	varchar	Campus where student was enrolled
Start_time	datetime	Start time (in year-month-date hh:mm:ss format) of each training session
End_time	datetime	End time (in year-month-date hh:mm:ss format) of each training session
Period	int	Duration (in minutes) of each training session.
Subject	varchar	Training topic
Trainer	varchar	Name of coach for each training
License	varchar	License type

Appendix C

Table C1. Definition of Variables

Variable	Definition
Testday	Interval in days between the date of last K2 training session and the K2 test pass date
Simulation	Dummy variable equals 1 if the student took a simulation test and 0 otherwise
Gapstd	Standard deviation of interval (in minutes) between each two training sessions
Gapmean	Mean of interval (in minutes) between two training sessions
Ptestday	Interval between the simulation test date and the K2 test pass date divided by the interval between the K2 test pass date and the date of the last training session
Feedbackday	Interval in days between simulation test date and date of the last training session
Falsenum_mean	Mean number of errors made by the student in the simulation tests.
Falsenumdummy	Dummy variable equals 1 if the student's Falsenum_mean is larger than the median of sample's Falsenum_mean
Age	Age of student
Female	Dummy variable equals 1 if the student is female
Trainday	Totals days between the date of last training session and the date of the first training session
Sub2period	The number of periods (in minutes) of each training session.

About the Authors

Wen Xie is a doctoral student in management information systems in the Department of Management & Marketing at Hong Kong Polytechnic University. Her research interests include artificial intelligence in business, IT innovation and consumers and digital platforms.

Xin Xu is an associate professor in the Department of Management & Marketing, Faculty of Business, at The Hong Kong Polytechnic University. He received his PhD in information systems from Hong Kong University of Science and Technology. His current research interests include IT innovation management, social media analytics, mobile computing, and human-computer interaction. His work has appeared in leading academic journals—*MIS Quarterly, Information Systems Research, Management Science, Information Systems Frontiers, IEEE Transactions on Engineering Managemen*t, and *Journal of the Association for Information Systems*.

Ruiqi Liu is a PhD student in the School of Accounting & Finance at Hong Kong Polytechnic University. She received her master's and bachelor's degrees in accounting from Beihang University, China. Her current research interests include entrepreneurial finance and entrepreneurial public markets, behavioral finance, corporate disclosure and corporate governance, and interdisciplinary research.

Yong (Jimmy) Jin is an assistant professor in the School of Accounting & Finance, Faculty of Business, at Hong Kong Polytechnic University. He received his PhD in business administration from the University of Florida. His current research interests include interdisciplinary research in information systems, technology management and finance. His work has appeared in leading journals such as *Production and Operations Management, Journal of the Association for Information Systems, Decision Support Systems, International Journal of Production Economics, Risk*, and others.

Qiang Li is a professor in the College of Economics & Management, South China Agricultural University and the director of Waterwood Capital Management. He received his doctorate in business administration from Hong Kong Polytechnic University. His current interests include internet of things, knowledge transfer, and venture capital and private equity investment.

Wenchao Bai is a professor in the College of Economics & Management, South China Agricultural University, and is also affliated with the Ho-sheng Law Firm. He received his doctorate in business administration from The Hong Kong Polytechnic University and is also the chair of the "One Thousand Talents" program in Sichuan, China. His current interests include internet of things, knowledge transfer and private equity investment.