When adaptive learning is effective learning: comparison of an adaptive learning system to teacher-led instruction

Shuai Wang , Claire Christensen , Wei Cui , Richard Tong , Louise Yarnall , Linda Shear & Mingyu Feng

Published online: 31 Aug 2020.

Submit your article to this journal ⃕

Article views: 55

View related articles ⃕

View Crossmark data ⃕

Routledge
Taylor & Francis Group

Check for updates

When adaptive learning is effective learning: comparison of an adaptive learning system to teacher-led instruction

Shuai Wang [a], Claire Christensen[a], Wei Cui[b], Richard Tong[b], Louise Yarnall[a], Linda Shear[a] and Mingyu Feng[c]

[a]Education Division, SRI International, Menlo Park, USA; [b]Squirrel AI Learning by Yixue Education Group, Shanghai, People's Republic of China; [c]Science, Technology, Engineering, & Mathematics Program, WestEd, San Francisco, USA

ABSTRACT

Adaptive learning systems personalize instruction to students' individual learning needs and abilities. Such systems have shown positive impacts on learning. Many schools in the United States have adopted adaptive learning systems, and the rate of adoption in China is accelerating, reaching almost 2 million unique users for one product alone in the past 3 years. Given such rapid adoption in China, it is useful to examine the efficacy of adaptive learning within that country's educational system. This study aimed to compare the learning impacts of individualized adaptive learning courseware to two common instructional approaches in China: large-group and small-group classroom instruction. This paper describes the results of two efficacy studies of one of China's first adaptive learning systems, Squirrel AI Learning. One study compares classroom-based individualized adaptive learning instruction to large-group instruction, and another to small-group instruction. Chinese eighth-grade students from two provinces randomly assigned to use Squirrel AI Learning showed greater gains on a mathematics test than those randomly assigned to whole-class or small-group instruction led by expert teachers. Findings provide a basis for further research into the selection, use, and impact of adaptive learning systems in Chinese education.

Introduction

Adaptive learning systems use various learning algorithms, such as artificial intelligence, machine learning, and item response theories to personalize the learning experience (e.g. Mavroudi et al., 2017; van der Linden, 2016). They are common in the United States and newer but gaining popularity in Asian countries such as China. As these systems become ubiquitous, parents, educators, and students must grapple with the trade-offs of in-person vs. adaptive instruction, particularly where in-person instruction is dangerous, difficult, or costly (e.g. pandemics, natural disasters, or remote locations). More research is needed to inform these decisions. While prior efficacy studies have shown positive effects of adaptive learning systems (VanLehn, 2011), more research is needed to disentangle the effects of adaptive learning, whole-class instruction, and small-group instruction. This paper reports two studies comparing the Squirrel AI Learning by Yixue Education Group mathematics intelligent learning system with expert teachers: one comparing Squirrel AI Learning classroom use to teachers teaching whole classes (20–30 students per class); the second comparing Squirrel AI Learning classroom use to teachers teaching small groups (3 students per class). In both contexts, students

---

CONTACT Shuai Wang ✉ shuai.derek.wang@gmail.com

using Squirrel AI Learning demonstrated significant learning gains over students receiving teacher instruction.

Literature review

Adaptive learning

For more than 3 decades, computer scientists and cognitive scientists have been developing adaptive learning systems that use artificial intelligence to mimic the interactions of human tutoring (Merrill et al., 1992). For example, developers have created these systems to present content, pose questions, assign tasks, provide hints, answer questions, or suggest improvements in the behavior or attitudes of learners (Ma et al., 2014). All adaptive learning systems follow a similar "closed loop" core architecture that gathers data from the learner and then uses those data to estimate the learner's progress, recommend learning activities, and provide tailored feedback. The adaptive system's algorithms typically make such decisions by referring to a domain model of the knowledge to be learned, a student model of learners' background characteristics (knowledge level, affect, and motivation), and a task model that specifies features of the learning activities (such as questions, tasks, quizzes, dynamic hints, feedback, prompts, and recommendations) (Lee & Park, 2008).

As access to computers and internet connectivity become more common inside and outside schools, interest in technology-based learning systems, including adaptive learning systems powered by artificial intelligence, has increased. Over the last 2 decades, public perception has become increasingly hopeful that artificial intelligence can improve education (Fast & Horvitz, 2017), and adaptive learning systems are growing increasingly common in the United States. The best-known systems include Knewton, ALEKS, i-Ready, DreamBox Learning, Achieve3000, and a variety of cognitive tutors.

Research, particularly in U.S. schools, shows that adaptive learning systems can promote student learning (e.g. Jones, 2018; VanLehn, 2011). Of 37 recent studies examining the effects of adaptive learning on learning achievements, 86% (32 studies) reported positive effects (Xie et al., 2019). In a comparison of 6,400 courses, 1,600 of which were adaptive, adaptive courses were found to improve student performance more effectively than non-adaptive courses (Bomash & Kish, 2015). Positive effects were reported in a large-scale effectiveness study of an adaptive learning system, Cognitive Tutor Algebra I, in 147 middle and high schools in seven U.S. states. After 2 years of the intervention, the average student's performance improved by approximately 8 percentile points (Pane et al., 2014). A more recent study found that over 2 years, personalized learning had positive effects on student mathematics and reading performance. Further, the lowest performing students made substantial gains relative to their peers (Pane et al., 2017). Yet more research is needed to understand the conditions under which adaptive learning is associated with the greatest gains – for example, relative to what learning conditions? Further, more research is needed on the effects of adaptive learning on students in middle and high school. A systematic review of articles on adaptive and personalized learning published in Social Sciences Citation Index journals from 2007 to 2017 found that of 70 articles, only 9% included middle and high school students, whereas 46% were conducted with higher education students (Xie et al., 2019).

The need for adaptive learning in China

While adaptive learning is in the early stages of development and adoption in China, the nation is primed for uptake. China has over 144 million online education users (China Internet Network Information Center, 2017). Adaptive learning is a stated Chinese education policy priority (e.g. O'Meara, 2019), perhaps because it has potential to address some of China's persistent educational challenges, such as inequity. While education equity in China has improved significantly, it is still below the international average (Zhang et al., 2007).

Adaptive learning may help to address the drawbacks of large Chinese class sizes, a potential source of inequity. Chinese junior secondary schools have an average class size of 52 students (Organisation for Economic Co-operation & Development, 2012a, 2012b) compared to average middle school class sizes of 16.7 students in the United States (U.S. Department of Education, 2011–12). In such large classes, it can be difficult for instructors to personalize learning to address individual students' needs, particularly for disadvantaged students. Large classes are associated with worse outcomes for low-income, minority, and struggling students than for other students (Bosworth, 2014; Schanzenbach, 2014). Adaptive learning may support equity by offering personalized instruction to every student, regardless of their ability level.

Chinese families are increasingly using private small-group or individual tutoring to address perceived shortcomings of large class sizes; however, a reduction in class size alone may not be sufficient to benefit students. According to China's 2005 Urban Household Education and Employment Survey, approximately two-thirds of junior high school students and half of senior high school students participated in private tutoring (Xue & Ding, 2008; cited in Zhang & Liu, 2016). While Asian students and their families may believe that one-to-one and small group tutoring boosts academic achievement (Zhan et al., 2013), research shows that the reduction in class size alone is insufficient to promote learning gains. In some studies, East Asian and East Asian American students fared worse in individual or small-group tutoring than in lecture-style instruction (Byun, 2014; Byun & Park, 2012; Zhang & Liu, 2016). Adaptive learning may be a new solution for ensuring rigor in students' tutoring experiences.

Use of adaptive learning systems may also help to address China's persistent rural-urban disparities in educational quality. A nationally representative survey found that rural Chinese students' math and Chinese-language scores are 2 years behind their urban peers (Wang et al., 2018). Compared to rural schools, urban schools in China have more teachers per student and more teachers with advanced degrees (National Education Inspection Team, 2009; as cited in Peng et al., 2014). Rural schools are less likely to have enough teachers with expertise in non-core subjects and more likely to mismatch teachers to subject areas to fill these gaps (National Education Inspection Team, 2009; as cited in Peng et al., 2014; Peng et al., 2014). Adaptive learning may help to ensure that rural students receive high-quality instruction, including in non-core subjects.

The Squirrel AI Learning system

Squirrel AI Learning is one of the first Chinese developers to introduce an adaptive learning system in its online education platform. Squirrel AI Learning has developed instructional materials for middle school mathematics, English, physics, Chinese, and chemistry (for a fuller description of the Squirrel AI Learning system, see Li et al., 2018). The Squirrel AI Learning system, since its development in 2016, has established over 2,000 learning centers in over 700 cities serving almost 2 million registered accounts. This usage base represents a broad range of student populations with respect to socioeconomic status, urbanicity, and academic achievement. Squirrel AI Learning's rapid uptake since its start in 2016 indicates that adaptive learning addresses substantial needs in the Chinese after-school tutoring market. More research is needed to ensure that this growing trend is truly supportive of students' learning in this context.

Unlike traditional Chinese teacher-centered instruction, the Squirrel AI Learning online system provides student-centered, personalized, interactive and data-analytics-driven instruction to enrich students' learning experiences. Figure 1 demonstrates the Squirrel AI Learning system. Squirrel AI Learning's product design is grounded in many of the same principles that inform excellent instructor-to-student instruction. These include:

- Formative assessments to determine the student's ability level,
- Problems targeted to the student's ability level,
- Instant, intelligent feedback including elaborated explanations, and
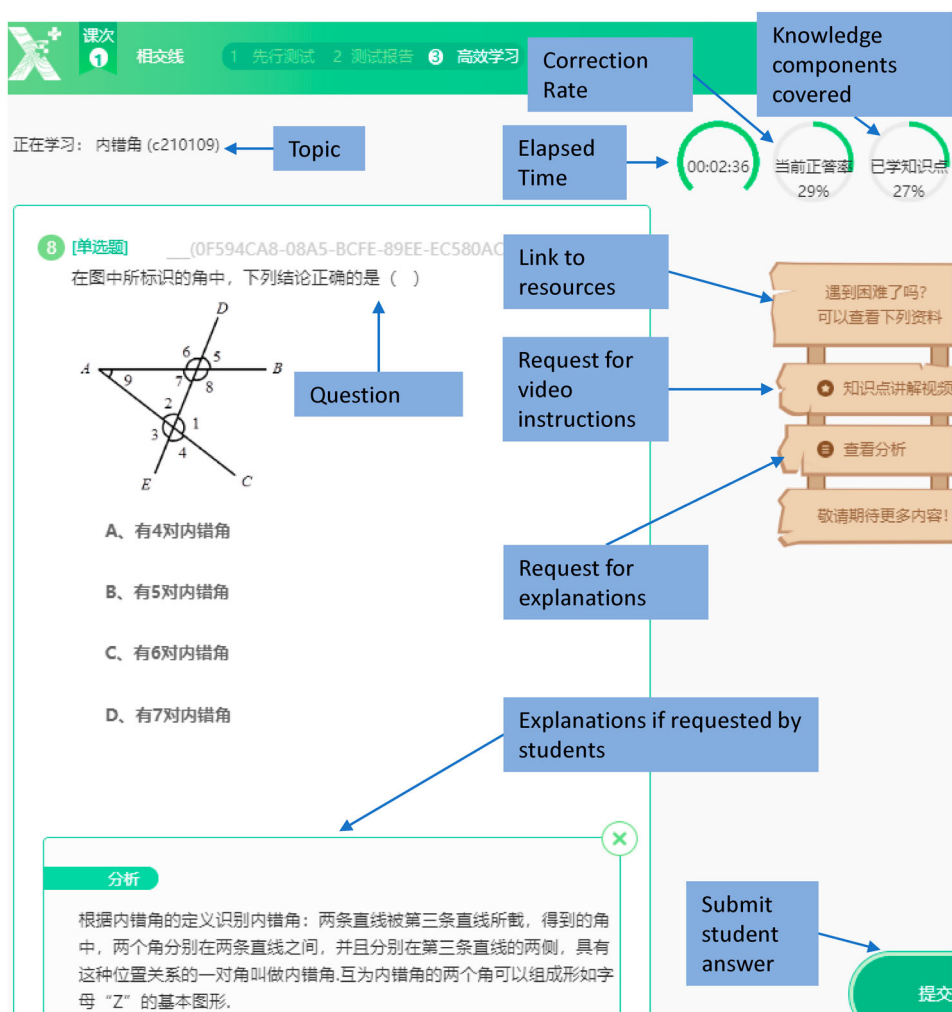- Supports (e.g. tutorials) differentiated by ability level.

Figure 1. Demonstration of Squirrel AI Learning adaptive learning system.

Research is needed to determine whether Squirrel AI Learning can address China's education challenges by improving upon traditional large-class instruction or small-group tutoring. Preliminary studies have found that students learn more from Squirrel AI Learning than from whole-classroom instruction (e.g. Li et al., 2018; Feng et al., 2018; Wang et al., 2019). More rigorous research with larger samples is needed to establish Squirrel AI Learning's benefits over large-group and small-group instruction. Such research is especially valuable as educators and families seek options for online education during closures related to the COVID-19 pandemic.

The present studies

This paper is one of the few to explore the promise of adaptive learning systems for educating Chinese students. Its goal is to inform decisions that Chinese educators, parents, and students will face increasingly as adaptive learning becomes more common in China. To inform those decisions, this paper compares the efficacy of adaptive learning systems to whole-class and small-group instruction. It explores two research questions:

(1) Does Squirrel AI Learning improve student learning, compared with expert teachers' **whole-class** instruction?
(2) Does Squirrel AI Learning improve student learning, compared with expert teachers' **small-group** instruction (e.g. class size = 3)?

Study 1

Sample and random assignment

In the first experiment, 200 eighth-grade students (ranging from 13 to 15 years old) were recruited from 20 middle schools in a major city of Sichuan Province. Students received school supplies (e.g., pens, ruler, and eraser) for participating in the study. The research team used a SAS random generator for random assignment: 100 students were randomly assigned to the treatment group, in which students used the Squirrel AI Learning system; the remaining 100 students were assigned to the control group. Students were not aware of their randomly assigned condition until they arrived at the study site on the first day of the experiment. Students were randomized to groups prior to the baseline assessment. We addressed baseline equivalence analytically, instead of creating baseline-equivalent groups using baseline assessment scores, becauscoring the baseline assessment and conducting the randomization would have delayed students' ability to begin instruction during the three-day experimental window, reducing overall instructional time.

Research procedure

For each of 3 days over a weekend and a national holiday, participating students engaged in math instruction at a designated school. The study's intensive approach was designed to maximize instructional time. Both groups participated in 5 h and 50 min of instructional time over the 3 days. Study topics included Pythagorean theorem and its applications, real numbers, solving triangles, and expressions. Both groups followed the same learning schedules, including study length, break length, and lunch time. Intervention implementation was monitored by an independent research organization.

Students assigned to the treatment group used computers individually in computer labs separate from the control students. Treatment students did not receive human instruction. Computer labs were overseen by a researcher who provided only technical assistance and schedule information. All Squirrel AI Learning use occurred during the scheduled study periods; students were not given access to the Squirrel AI Learning system outside of scheduled study hours. Prior use is not required for Squirrel AI Learning to adapt the learning experience – each time the student uses the system, the system assesses student ability and adapts accordingly.

Students assigned to the control group were arranged in classes of 20–30 and taught by one of three expert teachers. To ensure that the control group received quality instruction, teachers in the control group were all award-winning experienced teachers, with recognitions from the mathematical education field such as "District Excellent Teacher" and "Excellent Backbone Teacher." The teachers did not know the students before the experiment. Control students did not use Squirrel AI Learning during the experiment.

Data source

Two equivalent test forms were created prior to the study to serve as pre- and post-tests. Both forms covered standard content for grade 8 students, especially those topics covered during the instructional period (described above). These items were review for the students in the study, as is typical in after-school tutoring programs in China. Items in the two forms were constructed by an independent experienced math teacher in a local school, who was not part of the research team

or the study. Two independent content experts reviewed the two forms and adjusted the tests as necessary to ensure comparability of the forms. Each form included 20 items, including multiple choice, fill-in-the-blank, and short essays, with each form worth a total of 100 points. Prior to the study, the research team piloted the tests on a small group of students and performed statistical tests (including Cronbach's alpha and Item Response Theory) to ensure the reliability and validity of the two tests.

On the first day of the study, before the instruction started, the research team administered a paper-pencil pre-test to both groups. Students were randomly assigned to test forms such that half of the students in each group (treatment or control) received each test form. On day 3 at the end of the study they completed the post-test, using the opposite form they took at pre-test. All students finished both tests within the 60-minute testing window. Students also completed demographic surveys at the end of the study, including student age (numerical variable), gender (categorical variable), and parent education level (for the parent who supervised the study the most, collected as a proxy for social economic status, given that free and reduced price meals are not implemented in China). For analytic purposes we divided parent education level into four levels: graduate degree, college degree, high school degree, and below high school.

Analysis and results

At the end of the study, 90 out of 100 treatment students remained in the study, while 73 out of 100 control students remained. These attrition rates reflect students' personal schedules conflicting with the time of the experiment – a weekend and a national holiday. However, the substantial difference between the attrition rates (10% for treatment vs. 27% for control) could also be attributable to student attitudes towards different experimental conditions. Therefore, following U.S. Department of Education What Works Clearinghouse™ (WWC) Standards Handbook Version 4.0 (U.S. Department of Education, 2018) best research practices, the research team analyzed the data as a quasi-experiment, rather than analyzing the data as a randomized controlled trial.

We examined baseline equivalence to ensure group comparability, using key measures suggested by WWC. Analyses included only students who remained in the study, completed both pre-test and post-test, and provided complete responses to all survey questions in the analysis models. The final sample was 155 students: 87 treatment students and 68 control students. There was no significant difference between the treatment group pre-test scores ($M = 60.08$, $SD = 16.94$) and the control group pre-test scores ($M = 56.06$, $SD = 17.22$), $t(153) = 0.98$, $p = 0.33$, $g = 0.16$. Similarly, there was no significant difference in parent education level between the treatment group ($M = 3.03$, $SD = 1.25$) and the control group ($M = 2.87$, $SD = 1.28$), $t(153) = -0.82$, $p = 0.42$, $g = 0.13$. These findings indicate that treatment and control students were similar despite the differential attrition.

Next, we examined whether students in both groups improved from pre-test to post-test. In both groups combined, scores on the post-test were higher than pre-test, $t(154) = 5.72$, $p < .001$, $g = 0.31$, suggesting that math achievement improved significantly overall (see Table 1). For the treatment group, post-test scores were significantly higher than pre-test scores, $t(86) = 7.08$, $p < .001$, $g = 0.48$; however, for the control group, post-test scores were similar to pre-test scores, $t(67) = 1.03$, $p = .31$, $g = 0.08$.

Table 1. Pre-test, post-test, and improvement scores of students taught by expert human teachers and Squirrel AI Learning in study 1.

Test	n	Pre-test		Post-test		Improvement		
		M	SD	M	SD	M	t	g
Treatment	87	56.66	17.18	65.52	19.24	8.86	7.08***	0.54
Control	68	59.38	17.20	60.85	17.89	1.47	1.03	0.12
Combined	155	57.85	17.19	63.47	18.75	5.62	5.72***	0.46

Note: *** denotes $p < .001$ and g larger than 0.25 is considered substantial difference.

We conducted analysis of covariance (ANCOVA) to examine the differences between receiving instruction from Squirrel AI Learning and expert human teachers. We modeled post-test as the dependent variable, and included the following covariates: pre-test, as a student baseline measure; parent education level, as an indicator of socioeconomic status; and gender and age, as student background characteristics. We found that when we controlled for prior achievement and student demographics, students using Squirrel AI Learning significantly outperformed students taught by expert human teachers, $B = 7.00$, F (1, 149) = 13.99, $p < .001$, $R^2 = 0.64$. $g = 0.37$, with an associated improvement index of 14.4 percentile points, indicating that 64.4% of the students in the treatment group score above the mean of the control group, which is a substantial difference between treatment and control students. We further examined the interaction effects between group assignment (treatment or control) and prior achievement or student demographics, but no significant effects were identified.

Study 2

Whereas Study 1 compared Squirrel AI Learning to large-class instruction, Study 2 compares Squirrel AI Learning to small-group instruction. The sample, random assignment, research procedure, and data source were largely identical to Study 1. The following specifies differences from Study1.

Sample and random assignment

In the second experiment, 102 eighth-grade students (ranging from ages 13–15 years old) were recruited from four middle schools in a major city of Shandong Province. Students were randomly assigned to condition and not informed of their condition until they arrived on the first day.

Research procedure

The study lasted 3 days over a Friday and a weekend. Over these 3 days, both groups spent a total of 8 h and 30 min studying. Study 2 included more instructional hours than Study 1 for logistical reasons. Study 1 occurred over a national holiday. To minimize attrition, the Study 1 design minimized demands on students' holiday time. Study 2 occurred over a non-holiday Friday and a weekend, when we felt we could expose students to more instruction with less risk of attrition. Study topics included single-variable and multi-variable quadratic functions. Treatment group students again used Squirrel AI Learning on individual computers in computer labs. Comparison group students were taught by 17 award-winning expert teachers, with an average of 22 years of teaching experience, all of whom had received awards for "District Excellent Instructor." Each of the control expert teachers taught three students in their class, to allow us to compare the effects of Squirrel AI Learning to small group instruction.

Data source

Two equivalent test forms were created as in Study 1. Each form included 21 items, worth a total of 100 points. Both forms covered standard Grade 8 review content, especially those topics taught during instructional time (see above).

Analysis and results

All 51 treatment students and 51 control students completed the study. There was a significant difference in pre-test scores between the treatment group ($M = 77.22$, $SD = 11.26$) and the control group ($M = 81.37$, $SD = 10.12$), $t(100) = 1.96$, $p = 0.05$, $g = 0.39$. Thus, we performed propensity score matching (PSM) to identify a control group that was similar to the treatment group. PSM is a quasi-

Table 2. Pre-test, post-test, and improvement scores of students taught by expert human teachers and Squirrel AI Learning in study 2.

Test	N	Pre-test		Post-test		Improvement		
		M	SD	M	SD	M	t	g
Treatment	48	77.10	11.40	83.71	10.75	6.60	5.08***	0.73
Control	36	77.50	9.42	77.06	9.40	−0.4375	0.28	0.05
Combined	84	77.27	10.54	80.86	10.66	3.59	3.35***	0.36

Note: *** denotes $p < .001$ and g larger than 0.25 is considered substantial difference.

experimental statistical matching technique that creates artificially similar analytic samples to reduce possible statistical bias from observed variables. We employed nearest neighbor matching and allowed treatment students to match with multiple control students. We calculated PSM weights to account for the number of times a given control student was matched with treatment students. Students with key missing variable were deleted from the analytic sample. The matched sample included 48 treatment students and 36 control students. All inferential analyses below were conducted in the matched sample.

We examined whether the full matched sample improved from pre-test to post-test. Scores at post-test were higher than at pre-test, $t(83) = 3.35$, $p < .001$, $g = 0.36$, suggesting that math achievement improved significantly among both groups combined (see Table 2). For the treatment group, post-test scores were significantly higher than pre-test scores, $t(47) = 5.08$, $p < .001$, $g = 0.73$; however, for the comparison group, post-test scores were similar to pre-test scores, $t(35) = 0.28$, $p = .78$, $g = 0.05$.

After we controlled for student prior achievement and demographics, students using Squirrel AI Learning ($M = 83.71$, $SD = 10.75$) significantly outperformed the control group ($M = 77.06$, $SD = 9.4$), $B = 7.10$, $F(1, 78) = 15.37$, $p < .001$, $R^2 = 0.45$. $g = 0.69$, with an associated improvement index of 25.5 percentile points, indicating that 75.5% of students in the treatment group scored above the mean of the control group, a substantial difference. We further examined the interaction effects between group assignment (treatment or control) and prior achievement or student demographics, but no significant effects were identified.

Discussion

This paper adds to the research on the efficacy of adaptive learning systems for secondary mathematics instruction. It is among the first studies that examine the efficacy of adaptive learning systems in China, demonstrating that adaptive learning can be effective in Chinese contexts.

In both studies, students who used Squirrel AI Learning independently outperformed those taught by expert teachers, which is consistent with prior research in other regions (Jones, 2018). This suggests that adaptive systems effectively accommodate individual students' levels of knowledge and learning needs in ways that even highly skilled teachers in classroom settings do not. Learning gains did not differ based on students' prior knowledge, gender, age, or parental education, suggesting that Squirrel AI Learning adapts to students' needs to promote similar learning gains for all students. The combined results from these two studies indicate that the benefits of the Squirrel AI Learning system cannot be attributed to a reduction in class size alone, as students using Squirrel AI Learning outperformed students receiving both whole-class and small-group instruction. Adaptive learning systems have the capacity to mimic a one-on-one tutoring experience, a design feature that likely played a role in the results. In future studies, we hope to build on these findings by comparing Squirrel AI Learning to individual instruction from an expert human tutor. It is important to note that the effect sizes from the two studies should not be directly compared because the two studies covered different mathematical topics, included students from different provinces, had different durations, and included different control-group teachers. The purpose of this manuscript is not to directly compare the two studies, but rather to compare the treatment and control group within

each study. The treatment and control students in each study were exposed to the same duration and topics of instruction.

The duration of these two studies (3 days) supports this study's internal validity while perhaps limiting external validity. Most educational technology evaluations allow for students to use the product naturalistically over a longer period. In these studies, the treatment-control contrast may be muddied by differences in dosage or usage patterns. Using a condensed, tightly controlled study schedule improves the present study's internal validity by eliminating dosage as an alternative explanation. Future studies over a longer duration are needed to support the external validity of these claims. Perhaps human instructors are more skilled than Squirrel AI Learning at organizing their instruction over time. Perhaps learning gains decrease as the novelty of Squirrel AI Learning wears off.

Further research is warranted to examine the efficacy of the Squirrel AI Learning system for other grade levels, mathematical topics, and academic areas, and in comparison with other adaptive learning products. It would be useful to examine other learning outcomes in future studies, such as complex problem solving with real-world tasks and extended recall (delayed post-test). More research is needed to understand Squirrel AI Learning's efficacy for introducing new content; the present studies focused on selected math topics which were considered review for students at this grade level. Further, more research is needed to understand whether other adaptive learning products are effective in this context; the present studies were limited to the use of Squirrel AI Learning, which may use unique algorithms and features (Li et al., 2018).

Conclusion

Few experimental studies to date have evaluated the effectiveness of learning systems such as Squirrel AI Learning in China. However, as China's technology infrastructure develops and more schools introduce educational technology into classrooms, interest in this type of research will continue to grow. The present studies contribute much-needed knowledge to this field and provide a basis for further research into the selection, use, and impact of adaptive learning systems in Chinese classrooms.

Disclosure statement

Funding

Notes on contributors

Shuai Wang, Ph.D., an education researcher at SRI International (previously known as Stanford Research Institute), has extensive experience in developing and evaluating diverse digital STEM education approaches, and specializes in research planning and statistical modeling in evaluations of education interventions. Dr. Wang has served as Principal Investigator, Co-Principal Investigator, and senior personnel for a large number of research projects, including studies funded by U.S. National Science Foundation (NSF) and U.S. Department of Education. His STEM education work has led to many top-tier journal publications, book chapters, and conference presentations. These have resulted in global media coverage, including the U.S. National Science Foundation homepage.

Claire Christensen, Ph.D., is a senior education researcher at SRI International. She leads quantitative and mixed methods research on the use of media and technology to support children's learning.

Wei Cui, Ph.D., is a co-founder and Chief Scientist of Squirrel AI Learning by Yixue Education Group, the leading AI and adaptive education innovator for K-12 students in China. He led the development of Squirrel AI's intelligent adaptive learning system, and has published over 20 peer-reviewed academic papers and articles in AI, agent-based modeling,

complex adaptive system, quantitative finance and AI education. He was awarded MIT Technology Review "35 Innovators Under 35 China" in 2018.

Richard Tong, M.A., is the Chief Architect of Squirrel AI Learning by Yixue Education Group and the Chair of IEEE Learning Technology Standard Committee, 2020-2021. He is an experienced technologist, executive, entrepreneur and one of the leading evangelists for standardization effort for global education technology.

Louise Yarnall, Ph.D., is a senior research social scientist in SRI Education with expertise in the evaluation and design of adaptive learning products. She uses learning science to improve instruction around college and career readiness and to support research and development of new forms of workforce learning technologies.

Linda Shear, M.A., is the Director of Commercial and International Strategy in SRI Education. Shear's research and consulting work focuses on systemic educational change and powerful uses of ICT (information and communications technology) to support deeper learning for all students.

Mingyu Feng, Ph.D., is a Senior Research Associate with the Science, Technology, Engineering, & Mathematics (STEM) program at WestEd, and has strong expertise in the design, development and research of technology-supported learning systems, and rigorous evaluation of impact of educational technologies on learning. She is also an expert in educational data mining and learning analytics, with extensive experience in analyzing usage data from different kind of learning systems and leveraging analytics to monitor implementation fidelity.

ORCID

Shuai Wang ⓘ http://orcid.org/0000-0002-4983-9558

References

Bomash, I., & Kish, C. (2015). *The improvement index: Evaluating academic gains in college students using adaptive lessons*. Knewton.

Bosworth, R. (2014). Class size, class composition, and the distribution of student achievement. *Education Economics*, *22*(2), 141–165. https://doi.org/10.1080/09645292.2011.568698

Byun, S. Y. (2014). Shadow education and academic success in Republic of Korea. In Hyunjoon Park & Kyung-Keun Kim (Eds.), *Korean education in changing economic and demographic contexts* (pp. 39–58). Springer. https://doi.org/10.1007/978-981-4451-27-7_3

Byun, S. Y., & Park, H. (2012). The academic success of East Asian American youth: The role of shadow education. *Sociology of Education*, *85*(1), 40–60. https://doi.org/10.1177/0038040711417009

China Internet Network Information Center. (2017). *The 40th China statistical report on internet development*. http://www.cnnic.cn/hlwfzyj/hlwxzbg/hlwtjbg/201708/P020170807351923266213.pdf

Fast, E., & Horvitz, E. (2017, February). *Long-term trends in the public perception of artificial intelligence*. In *Proceedings of the thirty-first AAAI conference on artificial intelligence*, San Francisco, CA.

Feng, M., Cui, W., & Wang, S. (2018). Adaptive learning goes to China. In C. P. Rosé, R. Martínez-Maldonado, H. U. Hoppe, R. Luckin, M. Mavrikis, K. Porayska-Pomsta, B. McLaren &, B. du Bulay (Eds.), *Proceedings of the 19th International Conference on Artificial Intelligence in Education (AIED)* (Vol. 10948, pp. 89–93). Springer. https://doi.org/10.1007/978-3-319-93846-2_17

Jones, A. (2018). *Interpreting Knewton's 2017 student mastery results*. https://www.knewtonalta.com/mastery/interpreting-knewtons-2017-student-mastery-results/

Lee, J., & Park, O. (2008). Adaptive instructional systems. In J. M. Spector, M. D. Merrill, J. V. Merriënboer, & M. P. Driscoll (Eds.), *Handbook of research on educational communications and technology* (3rd ed., pp. 469–484). Lawrence Erlbaum Associates.

Li, H., Cui, W., Xu, Z., Zhu, Z., & Feng, M. (2018). Yixue adaptive learning system and Its promise on improving student learning. In B. McLaren, R. Reilly, S. Zvacek, & J. Uhomoibhi (Eds.), *Proceedings of the 10th international conference on computer supported education* (Vol. 1, pp. 45–52). SCITEPRESS.

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

Mavroudi, A., Giannakos, M., & Krogstie, J. (2017). Supporting adaptive learning pathways through the use of learning analytics: Developments, challenges and future opportunities. *Interactive Learning Environments*, *26*(2), 1996–2010. https://doi.org/10.1080/10494820.2017.1292531

Merrill, D. C., Reiser, B. J., Ranney, M., & Trafton, J. G. (1992). Effective tutoring techniques: A comparison of human tutors and intelligent tutoring systems. *Journal of the Learning Sciences*, *2*(3), 277–305. https://doi.org/10.1207/s15327809jls0203_2

National Education Inspection Team. (2009). Summary: 2008 report on national education inspection: Pay attention to teachers of compulsory education. *Research in Educational Development*, *1*, 1–5.

O'Meara, S. (2019). Will China lead the world in AI by 2030? *Nature*, *572*(7770), 427–428. https://doi.org/10.1038/d41586-019-02360-7

Organisation for Economic Co-operation and Development. (2012a). *Education at a glance 2012: OECD indicators*. https://www.oecd-ilibrary.org/education/education-at-a-glance-2012_eag-2012-en

Organisation for Economic Co-operation and Development. (2012b). *PISA 2012 results*. http://www.oecd.org/pisa/keyfindings/pisa-2012-results.htm

Pane, J. F., Griffin, B. A., McCaffrey, D. F., & Karam, R. (2014). Effectiveness of Cognitive Tutor Algebra I at scale. *Educational Evaluation and Policy Analysis*, *36*(2), 127–144. https://doi.org/10.3102/0162373713507480

Pane, J. F., Steiner, E. D., Baird, M. D., Hamilton, L. S., & Pane, J. D. (2017). *How does personalized learning affect student achievement?* RAND.

Peng, W. J., McNess, E., Thomas, S., Wu, X. R., Zhang, C., Li, J. Z., & Tian, H. S. (2014). Emerging perceptions of teacher quality and teacher development in China. *International Journal of Educational Development*, *34*, 77–89. https://doi.org/10.1016/j.ijedudev.2013.04.005

Schanzenbach, D. W. (2014). *Does class size matter?* National Education Policy Center.

U.S. Department of Education. (2018). What works clearinghouse™ standards handbook (Version 4.0) (pp. 130). https://ies.ed.gov/ncee/wwc/Docs/referenceresources/wwc_standards_handbook_v4.pdf

U.S. Department of Education, National Center for Education Statistics, Schools and Staffing Survey (SASS). (2011–12). *Public school teacher data file, 2011–12*. Department of Education, National Center for Education Statistics.

van der Linden, W. (2016). *Handbook of item response theory*. Chapman and Hall/CRC.

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

Wang, Shuai, Feng, Mingyu, Bienkowski, Marie, Christensen, Claire, & Cui, Wei. (2019). Learning from an adaptive learning system: Student profiling among middle school students. In H. Lane, S. Zvacek, & J. Uhomoibhi (Eds.), *Proceedings of the 11th International Conference on Computer Supported Education (CSEDU)* (Vol. 1, pp. 78–84). SCITEPRESS. https://doi.org/10.5220/0007729700780084

Wang, L., Min, W., Zhang, S., Shi, Y., & Rozelle, S. (2018). Math and Chinese-language learning: Where are China's vulnerable subpopulations? *Asian Survey*, *58*(5), 797–821. https://doi.org/10.1525/as.2018.58.5.797

Xie, H., Chu, H. C., Hwang, G. J., & Wang, C. C. (2019). Trends and development in technology-enhanced adaptive/personalized learning: A systematic review of journal publications from 2007 to 2017. *Computers & Education*, *103599*. https://doi.org/10.1016/j.compedu.2019.103599

Xue, H., & Ding, X. (2008). An empirical study on private tutoring for students in urban *China. Economics of Education Research (Beida)*, *6*(1), 1–14. (in Chinese).

Zhan, S., Bray, M., Wang, D., Lykins, C., & Kwo, O. (2013). The effectiveness of private tutoring: Students' perceptions in comparison with mainstream schooling in Hong Kong. *Asia Pacific Education Review*, *14*(4), 495–509. https://doi.org/10.1007/s12564-013-9276-7

Zhang, C., Huan, Z., & Li, H. (2007). An empirical study on education equity in China, 1978–2004. *Frontiers of Education in China*, *2*(4), 536–544. https://doi.org/10.1007/s11516-007-0040-9

Zhang, Y., & Liu, J. (2016). The effectiveness of private tutoring in China with a focus on class-size. *International Journal of Educational Development*, *46*, 35–42. https://doi.org/10.1016/j.ijedudev.2015.11.006