A Study of Adaptive Learning in Large Class Sizes and the Enabling Conditions for Student Self-Regulated Learning in the UAE

Xin Miao, Alef Education, United Arab Emirates
Pawan Kumar Mishra, Alef Education, United Arab Emirates
Samantha Monroe, Alef Education, United Arab Emirates
Richard John Brooker, Alef Education, United Arab Emirates

The European Conference on Education 2023
Official Conference Proceeding

Abstract
In 2018, a 60 student program was piloted to explore the potential of using an adaptive learning system in larger class sizes to mitigate issues such as a lack of qualified teachers and high teacher turnover rates in the UAE. This study sought to understand the impact of this program on student engagement and academic performance, as well as the enabling conditions needed for student self-regulated learning. Using data from over 12,700 students' exam results, as well as surveys from teachers and students, we examined the impact of this program using a propensity score matching technique. Results of the study showed that increasing the teacher-student ratio had no significant negative impact on student academic performance, and in some cases increased student engagement. However, enabling conditions for student self-regulated learning and teacher feedback on this project provided key insights that guides a more in-depth digitization of the UAE K12 public education system, which has important policy and practice implications.

Keywords: Propensity Score Matching, Self-Regulated Learning, Adaptive Learning Systems, Teacher-Student Ratio, UAE K12 Public Schools

Introduction and Literature Review

Within the education sector, some of the most pervasive issues in schools around the world are in regards to teachers: a shortage of qualified teachers (Ingersoll, 2003), teacher absenteeism (UNESCO IIEP, 2022), and high teacher turnover rates (Ingersoll, 2001; Kraft, Marinell, & Yee, 2016). United Arab Emirates (UAE) K12 public schools are not immune to these issues, as its large number of expatriate teachers leave it vulnerable to the high turnover rates seen in other expatriate-heavy industries in the country (Goe et al., 2020; Qasim, 2020). In addition, the pay and working conditions for teachers in the UAE are not always ideal, such as low salaries, lack of discretionary funding, and cultural and societal limitations in comparison to other fields, which can lead to a shortage of qualified and experienced educators willing to both work in the country and remain working here after they start (Alhumaid, 2021). A teacher shortage has been documented in the UAE as early as 2012, but the lack of qualified teachers has grown exponentially during the pandemic (Abdallah & Alriyami, 2022). Studies found that hundreds of teachers resign in the country every year, some even leaving before the end of their contract and the academic year (Murdock, 2022). Due to the teacher issues mentioned above and their related impact on student learning outcomes, researchers and education policy makers explore class size as a possible means to mitigating the problem.

Studying the effects of class size on student performance has been a popular topic of research to explore the optimum provision of instructional inputs (e.g. teacher-student ratio, digital learning tools) versus student learning gains in various contexts, and its cost-effectiveness. A key factor that drives this research effort is cost reduction, particularly teacher salaries. According to a Brookings study, increasing the pupil/teacher ratio in the U.S. by one student would save at least $12 billion per year in teacher salary costs alone (Whitehurst & Chingos, 2011). Depending on educational context, the results are quite mixed. According to the OECD PISA study (2018) on class size and student reading performance, B-S-J-Z China, is a typical example of larger class size (42 students per class on average) and higher average reading score, compared with the rest of the OECD PISA participating countries. Similar results found that different teaching practices reduce the importance of class size in East Asia (Stigler et al., 1982; Blatchford et al., 2016; as cited in Baker et. al., 2020). In Western Europe and North America, smaller class sizes are associated with better learning outcomes (Glass, 1982; Shin & Chung, 2009; as cited in Baker et. al., 2020). Altinok & Kingdom (2012) found that in the Middle East, North Africa, and sub-Saharan Africa smaller class size is associated with better learning outcomes (as cited in Baker et. al., 2020). Leuven and Oosterbeek compiled data from several class size studies and their findings determined that while research showed both positive and negative impacts on student performance from altering class sizes, most effects were quite small and not significantly different from zero (2018). Hence, it really depends on the specific educational context.

With educational technology making its way into classrooms, one such tool that has risen in prominence is adaptive learning systems (ALS). Through features like intelligent instructional design, personalized learning paths, adaptive testing, and real-time learning analytics, ALS provides students with educational materials and tasks in varying levels of difficulty, maximizing their chances of success (Zarkovic, 2020). OECD highlights the merits of adaptive learning systems and their ability to enhance student self-regulated learning (2021). However, they stress the importance of teachers when using ALS, emphasizing that "adaptive systems are not meant to replace teachers, but rather enhance their role," with well-trained teachers able to increase the impact of these systems on student academic

achievement (OECD, 2021). With the potential of adaptive learning systems to enhance teacher roles and to enhance student self-regulated learning, there are interests to explore how the provision of adaptive learning systems and teachers impact class size and student learning gains, potentially reducing teacher salary cost.

Research Context and the 60 Student Program

The Alef Platform is a student-centered adaptive learning system developed by Alef Education. In 2017, the platform was piloted in Abu Dhabi K12 public schools as part of the national initiative to digitize education systems and explore how digital learning can solve existing challenges in public education. Through bite-sized multimedia content, learning analytics, and student-centric features, the Alef Platform provides students with personalized, engaging learning experiences. By providing continuous feedback about what students are working on and how they are performing, teachers can support and intervene at the right point in a student's learning process. Learning analytics dashboards can be used by school leaders to monitor school performance, identify gaps for intervention, and support teachers' professional development. From Grade 5 to Grade 12, Alef Platform is used as the primary learning resource.

In the 2018-2019 academic year, a project called the 60 student program was piloted in a public all-girls school, by merging two Grade 7 classes into a single class of approximately 60 students. These students used the Alef Platform for English, Math, and Science, as it is assumed that the Alef Platform assists student self-regulated learning. The main objective is to investigate to what extent students could self-regulate their own learning using the Alef Platform in a larger class size, and how teachers could support effective teaching. Later on, the program was piloted in four other schools through both administrator introduction and school self-selection. There are a few criteria based on which schools piloted this program: 1) there are classrooms big enough to accommodate 60 students; 2) school leadership supports the program; 3) students are, on average, high achievers compared to same-aged peers at their school. Additionally, all pilot schools received implementation support on infrastructure and teacher professional development. Infrastructural support includes providing enough white boards for all students to see teacher instructions, seating arrangements, charging and laptop devices; teachers are provided with regular professional development on blended learning best practices using the Alef Platform, classroom management, and instructional strategies to work with larger class size.

Research Rationale

Finding the optimal teacher-student ratio within the Alef school context has significant implications for UAE K12 public schools, as it provides an opportunity to understand the efficacy of providing educational inputs efficiently (e.g. Alef Platform, teacher-student ratio) while maintaining student learning gains. The design of the adaptive learning system is built on the assumption that students work at their own pace, receive help and feedback from the system, and are therefore less dependent on the teacher; teachers, on the other hand, can focus on providing assistance to students who need intervention. Introducing the Alef adaptive learning system modifies student engagement with the learning task and with the teacher; as a result, teacher knowledge and skills must also be updated. Furthermore, an increase in student-teacher ratio in this context could create additional tasks for the teacher, such as ensuring accountability for learning and managing student behavior and engagement in a larger class size.

Research Goal and Questions

The purpose of this research is to examine the impact of teacher-student ratios on student engagement and academic performance, the enabling conditions for self-regulated learning, and to receive feedback from 60 student classroom teachers on their teaching practices within blended K12 classrooms in the UAE using the Alef Platform. This was formulated through the following research questions:

RQ 1: What is the impact of increasing teacher-student ratio on student engagement and academic performance on the Alef Platform?

RQ 2: In the 60-student classrooms, what are the enabling conditions for student self-regulated learning with the Alef Platform?

RQ 3: What are the 60 student project teachers' feedback on this project?

Methodology

Quasi-Experimental Evaluation: Propensity Score Matching

The 60 student program was conducted without any prior evaluation planning, and schools were assigned to the program through a combination of self-selection and administrator selection. As a result, students participating in the program and those not participating had significantly different background characteristics, which made assessing its impact and effectiveness challenging. This is very common in many educational systems, however, understanding what works and what does not work scientifically are still crucial for making policy and practice implications.

In order to understand the impact of increasing student-teacher ratio on student engagement and academic performance, we used a quasi-experimental evaluation design, propensity score matching (PSM) which is widely used for assessing a range of educational interventions and is among the most recommended (Morgan & Winship, 2007; Andrillon et. al., 2020). Additionally, PSM fits well with the context of this study since the Alef Platform provides a wealth of background characteristics on students that can be used to identify fitting counterfactuals.

This technique works by finding a counterfactual group of students (i.e., students who did not participate in the program) that is similar to the treatment group on a range of background characteristics. This similarity is measured by the propensity score, which corresponds to the probability that a given student would be assigned to the 60 student classroom program. Students in the treatment group are then matched to students in the control group with the closest propensity score. This allows for an apples-to-apples comparison of the impact of the program on two groups of students that are similar in all respects but for their participation in the 60 student classroom program.

Implementation of PSM

The analysis was conducted in Python using an implementation of PSM provided by IBM (Shimoni et al., 2019). First we created a set of features for each student and subject based on the background characteristics available from the Alef Platform.

We then trained a logistic regression model with participation in the 60 student program as the dependent variable and the set of student background characteristics as the independent variables. This resulted in a propensity score for each student, which corresponds to the probability that they would be assigned to the program based on their background characteristics.

We then applied PSM using nearest neighbor matching with a caliper width of 0.2. This means that treatment students were matched to the nearest control students with the closest propensity score, where the absolute difference in propensity scores was less than 0.2. This method has the advantage of being relatively simple to interpret while also providing a robustness check (i.e., one can see if results change if a different matching algorithm is used).

After applying PSM, we compared the distributions of treatment and control students on a range of background characteristics to check for balance between groups. The distribution of the propensity scores can be seen in Appendix 1.

The impact of the program was then measured by taking the difference in the outcomes metrics between the treatment and matched pseudo-control group. 95% confidence intervals were then calculated based on 100 bootstrap simulations.

Treatment and Pseudo-Control Groups

We looked at a total of 12,780 students from 97 UAE K-12 public schools. Among the 97 schools, 4 Abu Dhabi K12 public schools participated in the pilot with a total of 241 students. Table 1 shows the number of students in treatment and non-treatment groups.

Grade	Non-Treatment	Treatment
Grade 7	2909	52
Grade 8	2956	90
grade10	3378	50
Grade 11	3296	49
Total	12539	241

Table 1: Number of Students in the Treatment Group and Non-Treatment Group by Grade

To create a pseudo-control, 241 of the 12,539 non-treatment students were matched against the treatment group.

Student Quantitative Measurements

To ensure an unbiased analysis, our propensity model was trained using a variety of quantitative and qualitative measures to ensure pseudo-control was comparable to our treatment group. This is crucial in removing any confounding variables and ensuring our analysis is unbiased (Austin, 2011; Caliendo & Kopeinig, 2019).

Table 2 below shows the measures used, which were calculated using data from the Alef Platform as well as the exam data provided by the UAE Ministry of Education (MoE). They include both engagement and performance measures aligned with guidelines provided by the

MoE. Average time spent on the Alef Platform and number of MLO[1] completed are chosen as key indicators for student engagement. Summative assessment scores[2] per MLO and high-stakes end of term exams conducted by the local Ministry of Education are chosen as student academic performance indicators. The aforementioned data points were aggregated at the student, subject, and academic term levels. Aggregated features of the first term (academic year 2021-2022) were then fed into a logistic regression model to calculate the propensity scores, which were used to identify the pseudo-control group.

To answer RQ1, we measured the average effect of the treatment by comparing it to the outcomes of the pseudo-control on metrics from the second term (academic year 2021-2022).

Features Category	Features Sub Category	Features	Data Type	Description
Independent	Engagement	Average time spent in Term 1	Numerical (mins)	Average time spent on platform
		Completed lessons in Term 1	Numerical	Number of MLO (micro learning objective) completed on platform
	Performance	Summative Assessment Score in Term 1	Numerical	Assessment on platform
		Exam Score Term 1	Numerical	Final Exam
Dependent	Engagement	Average time spent in Term 2	Numerical (mins)	Average time spent on platform
		Completed lessons in Term 2	Numerical	Number of MLO (micro learning objective) completed on platform
	Performance	Summative Assessment Score in Term 2	Numerical	Assessment on platform
		Exam Score Term 2	Numerical	Final Exam
	Conditions for Student Self-Regulated Learning	Emotional regulation and task performance, Metacognition, Motivation, interest and enjoyment, self-efficacy reading, self-efficacy using Alef	Numerical	Self-system
		Teacher support classroom disciplinary environment, Teacher active engagement and support, Teacher support for differentiation, Teacher support for independent learning,	Numerical	Metacognitive strategies

[1] MLO, short for Mini Lesson Objectives, is the naming convention for each curriculum lesson on the Alef Platform. The Alef Platform provides Grade 5-Grade 12 UAE public school students with a primary curriculum for English, Math, Science, Arabic, Islamic Studies and Social Studies.
[2] Summative Assessments evaluate student learning at the end of each MLO.

	Gender	Categorical (Nominal)	Demographic
	Grade	Categorical (Ordinal)	Grade level of students

Table 2: Features for Student Engagement, Performance, and Conditions for Student Self-Regulated Learning

Student Qualitative Measurements

In order to answer RQ2 on the enabling conditions needed for students to exercise self-regulated learning with the Alef Platform, a student survey was developed and implemented. The survey was constructed based on literature on theories on self-regulated learning (SRL), our previous research findings, classroom observations, and teacher interviews.

The Zimmerman and Winnie models are the most cited frameworks on self-regulated learning (Panadero, 2017) and provide a framework for understanding cognitive, metacognitive, and emotional aspects of learning. In Zimmerman's cyclical phases model, students first analyze the task, set goals, develop motivational beliefs, plan task execution, and activate relevant learning strategies. Then, students exercise control strategies to monitor progress and keep themselves engaged. The final step of the process involves students self-reflection on progress, success, and failure. Winnie and Hadwin's theory has a strong metacognitive perspective that recognizes self-regulated students as active and managing their learning (Panadero, 2017). Our previous research in Abu Dhabi public schools indicates that student math academic performance is significantly influenced by metacognition, self-efficacy, and affective reactions to tasks (Miao et al., 2021).

Based on the above literature, previous findings, and school visits, we designed the survey to measure some key conditions to exercise SRL with the Alef Platform, including self-system factors (e.g. self-efficacy, motivation and interest, emotional regulation) and metacognition, in consideration of enlarged teacher student ratio and the use of adaptive learning systems (see conditions for student self-regulated learning in Table 2 above). For instance, self-efficacy using Alef and other digital tools (e.g. I feel comfortable using Alef and other digital tools to do learning tasks) and self-efficacy reading Alef lessons (e.g. I cannot completely understand the text in Alef lessons by just reading it once on my own) are designed to reflect student expectations and judgements about self and learning tasks with Alef and other digital tools offered. Five questions are designed to measure self-efficacy using Alef and other digital tools, five questions are designed to measure self-efficacy reading Alef lessons. Emotional regulation and task performance measure student capacity to regulate stress and the learning environment (e.g. I feel stressed when I have trouble understanding Alef lessons on my own), and six questions are designed. For metacognition, six items were designed to understand student perception of the usefulness of metacognitive strategies learning an Alef lesson (e.g. I use Alef assessments to monitor my progress.

In terms of teacher enabling support for self-regulated learning, survey measurements were constructed in consideration of literature on key roles that teachers assist students to become independent learners (Meyer et al., 2008), as well as field visit feedback. Therefore, the classroom disciplinary environment (e.g. the teacher has to wait a long time for students to quiet down) is designed to capture how often students feel their learning is hindered due to disciplinary problems, which could be worsened due to larger teacher-student ratios. Teacher

active engagement and support (e.g. the teacher asks questions that motivate students to participate actively in Alef classes) to understand how often their teachers have been actively engaged in guiding students to set learning goals, monitor learning, provide feedback and help when needed, and if students feel they have a very positive relationship with the teacher. Teacher support for differentiation and independent learning (e.g. the teacher assigns different tasks for different students) is designed to capture strategies that teachers use to scaffold, differentiate, and support independent learning. Five questions are designed to measure classroom disciplinary environment, six questions are designed to measure teacher active engagement and support, and six questions are designed to measure teacher support for differentiation and independent learning, using frequency scale items as "every lesson or almost every lesson, most lessons, some lessons, never or hardly ever."

The student survey was translated into Arabic and delivered bilingually through Survey Monkey. Approvals were obtained from school principals and local education governing agencies before implementation. 331 student responses were collected from both control and treatment groups across Grade 7, Grade 8, Grade 10, and Grade 11. All data collected follows the government student privacy protocol. Cronbach's alpha with item analysis was used to measure the reliability of all survey questions, which yields excellent internal consistency with a Cronbach alpha value of 0.88.

Teacher Feedback Survey

The teacher survey was designed to get pilot teacher feedback on best practices within the 60 student project, the main areas of difficulty and potential improvement within the project, how the use of the Alef Platform helps teachers manage larger class sizes, and teacher perception of self-regulated learning of students within the program. The survey was offered in English through Survey Monkey and anonymously completed by 24 teachers across all four schools participating in the 60 student program, with the responses obtained from teachers purely for the purpose of internal feedback.

Based on the objectives, survey items were designed to reflect four key areas of teacher practices: classroom management, lesson planning, instructional practices, and professional responsibility, taking into account Danielson's Framework for Teaching (1996, 2007, as cited in Isoré, 2009, p. 11) and UAE Ministry of Education Observation Framework (Emirates Schools Establishment, n.d.; UAE Ministry of Education, n.d.). These teacher practices were indicators of what teachers should know and do to exercise "good teaching," plus how enlarging teacher-student ratios and the use of adaptive learning systems potentially would have an impact on these practices. Additionally, survey items were also constructed to get teacher general feedback on the pilot program.

Classroom management pertains to teachers' ability to create a classroom environment and culture that is conducive for learning, clearly presenting tasks for learning, as well as their ability to successfully manage student behavior, all with the larger class size. An example of a classroom management question is "[To what extent do you] struggle to find time to work with students one-on-one during class." Lesson planning focuses on teachers' ability to effectively plan for lessons within the project setting, their knowledge of their content and their students, selecting appropriate instructional goals, designing and presenting coherent lessons, and assessing student learning. One lesson planning question is "[How often in your 60 student classroom did you use] Alef student data and learning analytics dashboard." Instructional practices is teachers' ability to clearly and effectively communicate lesson

objectives, both mastering the language of instruction and encouraging students to use it as well, the successful ability to differentiate for student needs, the use of assessments and being able to provide effective feedback to students on time, and ensuring student progress is aligned from curriculum standards. Instructional practices questions include "[How much do you agree that] Alef diagnostic assessments are helpful to gauge students' knowledge to allow them to be placed into level-appropriate instructional groupings." Professional responsibility includes self-reflection on their own teaching to determine areas in which professional development is needed, as well as the ability to successfully manage both in-person and online teaching, often concurrently. An example of a professional responsibility question is "[To what degree do you currently need professional development in] teaching cross-curricular skills (e.g. learning-to-learn)." Finally, general feedback asks after their own self-efficacy in regards to teaching, their ability to promote student use of ICT, and teachers' opinions on the success of the 60 student project. An example of a general feedback question is "[Do you agree that] the program has been successful in meeting its objectives."

Findings

Research Question One

Our study shows that class size has a statistically significant impact on student engagement, but not performance. Table 3 below shows the average treatment effect (ATE) on students across subjects across the 6 months of Term 2. In English, Islamic studies, and social science subjects, students in the treatment group were statistically significantly more engaged than those in the pseudo-control group. Students in the treatment group had a higher rate of successfully completing lessons and spent more time engaging with content on the Alef Platform. In math, treatment group students spent less time studying compared to control students, but the result was not significant.

On average, student performance in the treatment group was not significantly different from that of the control. In English, summative assessment and exam scores had average treatment effects of -0.38 and 0.94, but the p-values were above 0.05 and not significant. Islamic studies, math, and social studies showed similarly non-significant results. Science was the exception: the summative assessment score was significant (p-value 0.01), with an average treatment score of -5.6.

Subject	Features	Average Treatment Effect (ATE)	P-Value	Confidence Interval (0.025, 0.975)	
English	Exam Score Term 2	0.94	0.37	-1.12	3.00
English	Summative Assessment Score in Term 2	-0.38	0.67	-2.13	1.38
English	Completed lessons in Term 2	10.39	*0.0001	6.23	14.56
English	Average time spent in Term 2	232.82	*0.0001	186.09	279.56
Islamic Studies	Exam Score Term 2	1.11	0.19	-0.57	2.80
Islamic Studies	Summative Assessment Score in Term 2	0.29	0.84	-2.53	3.10
Islamic Studies	Completed lessons in Term 2	1.62	*0.01	0.39	2.85
Islamic Studies	Average time spent in Term 2	8.36	0.65	-27.79	44.51

Math	Exam Score Term 2	0.76	0.56	-1.81	3.33
Math	Summative Assessment Score in Term 2	0.21	0.89	-2.76	3.19
Math	Completed lessons in Term 2	9.41	*0.0001	4.43	14.40
Math	Average time spent in Term 2	-93.06	0.08	-198.75	12.64
Science	Exam Score Term 2	1.57	0.38	-1.95	5.09
Science	Summative Assessment Score in Term 2	-5.60	*0.01	-9.69	-1.52
Science	Completed lessons in Term 2	5.21	0.12	-1.36	11.78
Science	Average time spent in Term 2	319.82	*0.0001	180.51	459.13
Social Studies	Exam Score Term 2	-0.52	0.44	-1.84	0.79
Social Studies	Summative Assessment Score in Term 2	0.30	0.83	-2.39	2.99
Social Studies	Completed lessons in Term 2	3.77	*0.0001	2.19	5.35
Social Studies	Average time spent in Term 2	80.08	*0.0001	39.52	120.64

Note: * mark shows significant p-value (P-Value < 0.05)

Table 3: Average Treatment Effect of Intervention by Subject on Measured Features

Research Question Two

To understand how increasing teacher-student ratios affect student internal conditions and teacher external support needed for students to exercise self-regulated learning using the Alef Platform, we applied the same PSM to our survey responses. We took 82 treatment group students who had answered our survey, and matched them against 85 similar non-treatment group students to create a pseudo-control group.

The results in Table 4 below show that student-perceived usefulness of metacognitive strategies were statistically significantly higher for students participating the 60 student program. Students in the treatment group reported exercising metacognitive strategies as particularly useful, as metacognitive strategies are core self-regulated learning skills that learners benefit from even in an adaptive learning environment.

Students in the treatment group also reported more frequent need of teacher support for independent learning and teacher active engagement and support than students in the control group. These two enabling conditions are key for a successful self-regulated learning environment in a larger class size. This makes sense as by increasing the teacher-student ratio, teachers might assign more independent learning tasks for students. However, larger class sizes need more frequent teacher support to be successful.

Features Category	Features	Average Treatment Effect (ATE)	P-Value	Confidence Interval (0.025, 0.975)	
Student internal conditions	Emotional regulation and task performance	0.009	0.796	-0.06	0.08
	Perceived usefulness of metacognitive strategies	0.166	*0.032	0.01	0.32
	Motivation, interest and enjoyment	-0.084	0.007	-0.14	-0.02

	Self-efficacy in reading	0.122	0.071	-0.01	0.26
	Self-efficacy using Alef	-0.051	0.452	-0.18	0.08
Teacher external support	Teacher Support on classroom disciplinary environment	0.046	0.359	-0.05	0.14
	Teacher active engagement and support	0.233	*0.002	0.09	0.38
	Teacher support for differentiation	0.068	0.312	-0.06	0.20
	Teacher support for independent learning	0.220	*0.002	0.08	0.36

Note: * mark shows significant p-value (P-Value < 0.05)

Table 4: Average Treatment Effect of Intervention on Measured Survey Responses

Research Question Three

In the teacher feedback survey, we found three of the four key focus areas of the survey yielded positive findings. The majority of interviewed teachers frequently use Alef Platform tools for classroom management; the Team feature allows teachers to group students and the Star-awarding feature motivates students on good behavior and improved performance. It was also claimed that Alef lessons reduced teachers' lesson planning time and allowed them to spend more time gaining a deeper understanding of the lesson beforehand. The Alef learning analytics dashboard was highly cited as beneficial to teachers in the 60 student program, particularly with regards to differentiation and real-time updates on student learning. Additionally, the vast majority of interviewed teachers report they encourage students to practice self-regulated learning in every or almost every lesson, and about half of them say the feedback provided directly to students by the Alef Platform helps facilitate this practice.

On the other hand, a large percentage of surveyed teachers said they struggled to provide subjective real-time feedback to students due to the large class size, and teachers regularly supplement Alef products to fill gaps in the Alef Platform, such as Google Classroom and ClassKick for providing subjective feedback and teacher-student real-time interaction. In order to improve accountability for learning in larger class sizes, features that allow teachers to have real-time interaction and written feedback on assigned learning tasks could be considered for product development.

The professional development section of the survey brought forward several areas in which teachers identified they need assistance, including how to better utilize Alef data and platform tools. With hopefully further improvements to the Alef Platform forthcoming, providing training for teachers and students on effective interaction and accountability features on the platform can help boost teacher use in these areas. Additionally, teachers voiced a need for professional development in teaching cross-curricular skills, such as problem solving, learning-to-learn, and independent learning, in which improving teacher efficacy in this domain would benefit all learners, but especially those in the 60 student program. Finally, assistance was requested for UAE-context specific needs, such as second language teaching. In UAE K12 public schools, math and science are taught in English. With students speaking Arabic as their first language, this means that teachers whose first language is Arabic would require professional development for teaching in English to better support these students both

in and out of the 60 student program. Therefore, while results showcase numerous positive aspects of the 60 student program, the noted areas of improvement are not insignificant and highlight the need for further assistance for teachers and potentially updates to the Alef Platform to ensure success within the program.

Conclusion

This study set out to evaluate the efficacy of increasing class size with the Alef adaptive learning system in the UAE K12 public school context using propensity score matching. Understanding the program impact on student engagement, academic performance, and enabling conditions for SRL, as well as teacher feedback on the program, provides important education policy implications and product improvement insights.

We found that class size has a statistically significant impact on student engagement, but not academic performance. Given that academic performance is a key metric for scaling up the project and this evaluation only tracks performance changes over 6 months, further research needs to be conducted to track program impact on academic performance particularly.

Among conditions to enable students to practice SRL, student-perceived usefulness of metacognitive strategies was statistically significantly higher for students participating in the treatment group. This means that students in larger class sizes believe their learning benefits from being able to use metacognitive strategies while interacting with the Alef adaptive learning system. Additionally, students in the treatment group reported more frequent support received for independent learning and active engagement from teachers than students in the control group, among teacher external support factors. Hence, providing students with these conditional support is a key consideration for improving student learning experience in larger class sizes.

Lastly, teachers highlighted several areas for improvement, such as the need for real-time interaction and accountability for learning features that allow them to see how student groups have been proceeding with learning tasks and to provide written feedback. Teachers also showcased numerous positive experiences, with the majority of teachers finding the Alef product features helpful in their daily practices on lesson planning, classroom management, and instructions. Additionally, learning analytics has been highly cited as a beneficial instructional feature that automates assessment and feedback real-time.

Implications and Recommendations

The Alef adaptive learning system provides the possibility to mitigate teacher-related challenges in the UAE K12 public school context, only if certain conditions are met based on our study.

Firstly, the design and development of the technological inputs (i.e. Alef adaptive learning system) requires an iterative improvement process, taking into account evidence-based user feedback. For instance, in our study, SRL requires students to learn independently, but learning itself is also a social activity and requires accountability. The Alef Platform, when used in larger class sizes, instead of only providing independent adaptive learning support and features that group learners together, real-time interactive features need to be designed to allow teachers and learner peers to share inputs, feedback real-time, which in turn supports engagement and accountability of learning.

Second, the ability to exercise metacognitive skills is a key 21st century skill to be taught explicitly in the UAE public K12 context, which has been proven important consistently in this and past studies. Being able to exercise metacognitive skills is recommended as a national education policy to ensure efficiency of learning in the trend of digitizing education systems, and teachers should be aware of the importance of such skills and should be trained to cultivate learner metacognitive skills.

Lastly, from teacher knowledge and skill perspectives, apart from PD on learner metacognitive skill acquisition, training teachers to use adaptive learning systems has a learning curve, especially if teachers are to use technology to benefit flexible teaching and learning and to support various learning tasks. Education systems should be ready to have skilled trainers and capacity to deliver this service. Teacher hiring and in-service training should take into account the fact that math and science subjects require second language teaching. AI-powered adaptive learning systems won't be able to replace teachers, yet teacher practices, knowledge, and skills will be modified.

Limitations

Given that the pilot was implemented in all-girls schools in the Abu Dhabi emirate, the findings may not be generalizable to other contexts (e.g. gender, other emirates with a different learning context). In addition, since the impact of the program was measured only over six months, it is difficult to determine its long-term impact, especially when it comes to student academic performance.

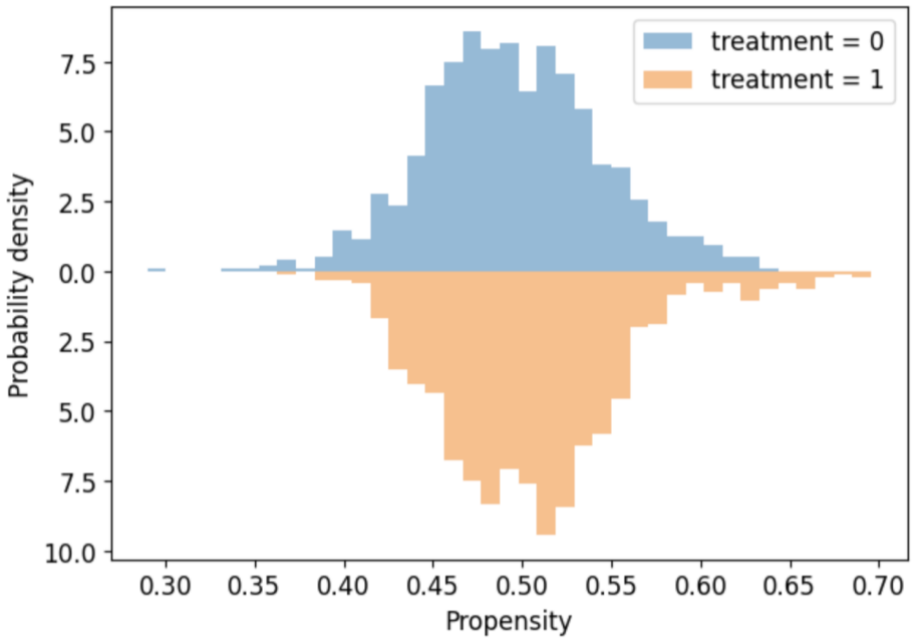
Despite the use of propensity score matching to control for treatment bias, important confounding factors and contexts couldn't be controlled. Additionally, nearly all of the students in our sample had already used Alef before starting treatment; if the program were started from scratch, results may have been different.

In spite of these limitations, this study provides valuable insights to guide decision-makers on possibilities of scaling up the project, conditions needed from students and teachers to better work with adaptive learning systems in the local educational context. Most importantly, increasing class size is doable if teachers and students have the conditions met to work with adaptive learning systems in the UAE context, and teacher feedback should be always incorporated for product improvement; however, further research is still needed.

Acknowledgements

The researchers would like to thank Alef Education for supporting this research project.

Appendix 1: Distribution of propensity score by treatment and pseudo-control group



References

Abdallah, A. K., & Alriyami, R. (2022). Changes in the education landscape caused by COVID-19: Opportunities and challenges from UAE perspective. *World Journal on Educational Technology: Current Issues*, *14*(3), 544–559. https://doi.org/10.18844/wjet.v14i3.7193

Alhumaid, K. (2021). The View of Emirati Students on Education as a Specialty: Understanding the Unwillingness to Join Faculties of Education in the UAE. *Journal of Management Information and Decision Sciences*, *24*(1S). Retrieved from https://www.abacademies.org/articles/the-view-of-emirati-students-on-education-as-a-specialty-understanding-the-unwillingness-to-join-faculties-of-education-in-the-uae-12511.html

Andrillon, A., Pirracchio, R., & Chevret, S. (2020). Performance of propensity score matching to estimate causal effects in small samples. *Statistical Methods in Medical Research*, *29*(3), 644–658. https://doi.org/10.1177/0962280219887196

Austin, P. C. (2011). An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behavioral Research*, *46*(3), 399–424. https://doi.org/10.1080/00273171.2011.568786

Baker, R. S., Al Yammahi, A., El Sebaaly, J., Nadaf, A., Kapp, T., & Adjei, S. (2020). Can Computer-Based Learning Environments Mitigate Large Class Size? *Conference: 28th International Conference on Computers in Education*. Retrieved from https://www.researchgate.net/publication/344363767_Can_Computer-Based_Learning_Environments_Mitigate_Large_Class_Size

Barrett, P., Treves, A., Shmis, T., Ambasz, D., & Ustinova, M. (2019). The impact of school infrastructure on learning: A synthesis of the evidence. *International Development in Focus*. https://doi.org/10.1596/978-1-4648-1378-8

Caliendo, M., & Kopeinig, S. (2008). Some practical guidance for the implementation of propensity score matching. *Journal of Economic Surveys*, *22*(1), 31–72. https://doi.org/10.1111/j.1467-6419.2007.00527.x

Eau, G., Hoodin, D., & Musaddiq, T. (2021). Testing the effects of adaptive learning courseware on student performance: An experimental approach. *Southern Economic Journal*, *88*(3), 1086–1118. https://doi.org/10.1002/soej.12547

Emirates Schools Establishment. (n.d.). Emirates Schools Establishment Formal Observation Form. Abu Dhabi, UAE.

García, E., & Weiss, E. (2019, March 26). *The teacher shortage is real, large and growing, and worse than we thought: The first report in 'The perfect storm in the teacher labor market' series*. Retrieved from https://www.epi.org/publication/the-teacher-shortage-is-real-large-and-growing-and-worse-than-we-thought-the-first-report-in-the-perfect-storm-in-the-teacher-labor-market-series/

Goe, L., Alkaabi, A. K., & Tannenbaum, R. J. (2020). Listening to and supporting teachers in the United Arab Emirates: Promoting educational success for the nation. *ETS Research Report Series*, *2020*(1), 1–18. https://doi.org/10.1002/ets2.12289

Ikpa, V. W. (2016). Politics, Adequacy, and Education Funding. *Education*, *136*(4), 468–472. Retrieved from https://www.ingentaconnect.com/content/prin/ed/2016/00000136/00000004/art00008#Refs

Ingersoll, R. M. (2001). Teacher turnover and teacher shortages: An organizational analysis. *CPRE Research Reports*, *38*(3), 3–35. https://doi.org/10.3102/00028312038003499

Ingersoll, R. M. (2003). The Teacher Shortage: Myth or Reality? *Educational Horizons: The Changing Characteristics of the Public Teaching Force, 81*(3), 146–152. Retrieved from https://www.jstor.org/stable/42926477

Isoré, M. (2009). (working paper). Teacher Evaluation: Current Practices in OECD Countries and a Literature Review. OECD Education Working Papers. Retrieved from https://www.oecd-ilibrary.org/docserver/223283631428.pdf?expires=1666264550&id=id&accname=guest&checksum=32CF78895036DAFAF8B215C298DA0504

Kraft, M. A., Marinell, W. H., & Shen-Wei Yee, D. (2016). School organizational contexts, teacher turnover, and student achievement. *American Educational Research Journal*, *53*(5), 1411–1449. https://doi.org/10.3102/0002831216667478

Leuven, E., & Oosterbeek, H. (2018, April). *Class size and student outcomes in Europe*. EENEE Policy Brief 2/2018. Retrieved from https://eenee.eu/wp-content/uploads/2021/05/PolicyBrief2-2018-3.pdf

Maffea, J. (2020). Lack of Resources in Classrooms. *English Department: Research for Change - Wicked Problems in Our World*. Retrieved from https://research.library.kutztown.edu/wickedproblems/38

Meyer, B., Haywood, N., Sachdev, D., & Faraday, S. (2008). (rep.). What is independent learning and what are the benefits for students? London: Department for Children, Schools and Families. Retrieved from http://www.curee.co.uk/files/publication/%5Bsite-timestamp%5D/Whatisindependentlearningandwhatarethebenefits.pdf

Miao, X., Mishra, P. K., & Nadaf, A. (2021). Evidence and promises of AI predictions to understand student approaches to math learning in Abu Dhabi K12 public schools. Gulf Education and Social Policy Review (GESPR), 1(2), 109–134. https://doi.org/10.18502/gespr.v1i2.8458

Morgan, S. L., & Winship, C. (2007). *Counterfactuals and Causal Inference: Methods and Principles for Social Research*. Cambridge University Press. Retrieved from https://edisciplinas.usp.br/pluginfile.php/3984640/mod_resource/content/2/%5BStephen_L._Morgan%2C_Christopher_Winship%5D_Counterfa%28BookFi.org%29%20%281%29%281%29.pdf

Murdock, A. (2022). Teacher retention in a dynamic cultural setting: A collective case study of expatriate teachers in the United Arab Emirates. *Handbook of Research on Teacher Education*, 29–57. https://doi.org/10.1007/978-981-19-2400-2_3

Nguyen, T. D., & Springer, M. G. (2019, December 4). *Reviewing the evidence on teacher attrition and retention*. Brookings. Retrieved from https://www.brookings.edu/blog/brown-center-chalkboard/2019/12/04/reviewing-the-evidence-on-teacher-attrition-and-retention/

OECD, & Baker, R. S. (2021). Artificial intelligence in education: Bringing it all together. In *OECD Digital Education Outlook 2021 pushing the Frontiers with Artificial Intelligence, Blockchain and Robots* (pp. 43–54). essay, OECD Publishing. Retrieved from https://learninganalytics.upenn.edu/ryanbaker/oecd-baker.pdf

OECD. (2009). Teaching Practices, Teachers' Beliefs and Attitudes. In *Creating Effective Teaching and Learning Environments: First Results from TALIS* (pp. 87–135). essay, OECD. Retrieved from https://www.oecd.org/berlin/43541655.pdf

OECD. (2018). Class size and mean reading performance. Retrieved from: https://www.oecd-ilibrary.org/education/class-size-and-mean-reading-performance_4ddb297c-en

OECD. (2021, September 28). *Teacher Practices*. Review Education Policies - Education GPS. Retrieved from https://gpseducation.oecd.org/revieweducationpolicies/#!node=41713&filter=all

Panadero, E. (2017). A Review of Self-regulated Learning: Six Models and Four Directions for Research. Frontiers in Psychology, 8:422. https://doi.org/10.3389/fpsyg.2017.00422

PISA. (2017, October). *Teacher Questionnaire for PISA 2018 General Teacher (International Option) Main Survey Version*. Retrieved from https://www.oecd.org/pisa/data/2018database/CY7_201710_QST_MS_TCQ-G_NoNotes_final.pdf

Pugliese, L. (2016, October 17). *Adaptive Learning Systems: Surviving the Storm*. EDUCAUSE Review. Retrieved from https://er.educause.edu/articles/2016/10/adaptive-learning-systems-surviving-the-storm

Qasim, A. (2020). The Impacts of High Employee Turnover in the UAE Hospitality Industry. *Human Capital in the Middle East*, 73–91. https://doi.org/10.1007/978-3-030-42211-0_4

UAE Ministry of Education. (n.d.). Ministry of Education Lesson Observation Form. Abu Dhabi, UAE.

UNESCO IIEP. (2022, July 29). *Teacher absenteeism*. IIEP Policy Toolbox. Retrieved from https://policytoolbox.iiep.unesco.org/policy-option/teacher-absenteeism/

UNESCO. (2020, July 7). *Distance learning solutions*. UNESCO. Retrieved from https://en.unesco.org/covid19/educationresponse/solutions

United Arab Emirates Ministry of Education. (2022, May 8). Digital Transformation Journey in Education. United Arab Emirates Ministry of Education. Retrieved October 20, 2022, from https://www.moe.gov.ae/En/ImportantLinks/Pages/DigitalTransformation.aspx

United Arab Emirates Ministry of Education. (2022, June 1). *School Inspection Framework*. Retrieved from https://www.moe.gov.ae/En/ImportantLinks/Inspection/Pages/EvaluationandQuality.aspx

Viac, C., & Fraser, P. (2020). (working paper). *Teachers' Well-Being: A Framework For Data Collection And Analysis*. OECD Education Working Papers. Retrieved from https://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=EDU/WKP(2020)1&docLanguage=En

White, H., & Sabarwal, S. (2014, September). *Quasi-experimental design and methods - unicef-irc.org*. Retrieved from https://www.unicef-irc.org/KM/IE/img/downloads/Quasi-Experimental_Design_and_Methods_ENG.pdf

Whitehurst, G.J. & Chingos, M.M. (2011, May 11). Class size: what research says and what it means for state policy. Brown Center on Education Policy at Brookings. Retrieved from https://www.brookings.edu/wp-content/uploads/2016/06/0511_class_size_whitehurst_chingos.pdf

Wilichowski, T., & Cobo, C. (2021, January 6). *Considering an adaptive learning system? A roadmap for policymakers*. World Bank Blogs. Retrieved from https://blogs.worldbank.org/education/considering-adaptive-learning-system-roadmap-policymakers

Zarkovic, S. (2020). Adaptive Learning Systems in Modern Classrooms. In *Ethical Use of Technology in Digital Learning Environments: Graduate Student Perspectives* (pp. 74–86). essay, University of Calgary. Retrieved from https://openeducationalberta.ca/educationaltechnologyethics/

Contact email: samantha.monroe@alefeducation.com