Exploring the Impacts of Implementing Block Mode of Teaching in Higher Education

Trinh, Nguyen Thi Thao, Ghapanchi, Amir and Purarjomandlangrudi, Afrooz (2024) Exploring the Impacts of Implementing Block Mode of Teaching in Higher Education. International Journal of Emerging Technologies in Learning (iJET), 19 (6). pp. 4-18. ISSN 1863-0383

06.2024

iJET

International Journal:
Emerging Technologies in Learning

Papers

Table of Contents

Papers

Online-Journals.org

iJET International Journal of Emerging Technologies in Learning

iJET | eISSN: 1863-0383 | Vol. 19 No. 6 (2024) | OPEN ACCESS

https://doi.org/10.3991/ijet.v19i06.49229

PAPER

Exploring the Impacts of Implementing Block Mode of Teaching in Higher Education

Nguyen Thi Thao Trinh[1], Amir Hossein Ghapanchi[1,2](✉), Afrooz Purarjomandlangrudi[3]

[1]College of Sport, Health and Engineering, Victoria University, Melbourne, Australia

[2]The Institute for Sustainable Industries and Liveable Cities, Victoria University, Melbourne, Australia

[3]College of Arts, Business, Law, Education and IT, Victoria University, Melbourne, Australia

amir.ghapanchi@vu.edu.au

ABSTRACT

The tertiary sector is one of the most competitive industries that makes significant contributions to countries' economies. Due to rapid changes in society, universities are urged to make suitable changes to attract students. One of the innovations some universities have adopted is changing their teaching mode to facilitate the learning process and attract students who find it more suitable for their learning style. The block model of teaching (BMT), a relatively new teaching mode in the higher education industry, has demonstrated great potential for increasing students' performance. This study employs a case study method and collects data from five institutes that have implemented BMT, namely: Salford University, Colorado College, Heriot-Watt University, Victoria University, and Quest University. The thematic analysis of the collected data has led to the proposal of a taxonomy comprising 14 positive impacts and seven negative impacts of implementing BMT, grouped into eight categories. Implications of the findings for studies and practitioners are also provided herein.

KEYWORDS

block model of teaching (BMT), intensive mode of delivery, accelerated teaching mode, impacts, outcomes, consequences

1 INTRODUCTION

The block model of teaching (BMT) is defined by Cawelti [1] as a teaching method that delivers the daily schedule or the daily amount of schooling in larger blocks of time. According to Murray, Barkat, & Pearlman [2], BMT can offer various benefits to students, making it an appealing teaching approach for universities. Cawelti [1] asserts that this method has been widely applied in secondary and high schools. This is because it provides high school students with more flexibility, allowing them to receive ample instructions from teachers to support their learning. Although the BMT is not a new concept for secondary or high school educators, its implementation

Trinh, N.T.T., Ghapanchi, A.H., Purarjomandlangrudi, A. (2024). Exploring the Impacts of Implementing Block Mode of Teaching in Higher Education. *International Journal of Emerging Technologies in Learning (iJET)*, 19(6), pp. 4–18. https://doi.org/10.3991/ijet.v19i06.49229

Article submitted 2024-02-21. Revision uploaded 2024-04-19. Final acceptance 2024-04-19.

at the tertiary education level is still limited. Furthermore, the reasons why some universities incorporate this teaching model into their programs are not fully understood. Studies on the impacts of BMT implementation in the higher education sector are very limited.

The study of the consequences of implementing BMT in the tertiary sector is quite limited. Some researchers have reported positive outcomes from BMT implementation. For example, Tatum [3] found that students who attended BMT classes tend to have better learning outcomes. Karjanto and Yong [4] discovered that BMT provides students and teachers with the opportunity to build relationships and allows teachers to closely monitor students' progress. On the other hand, some studies have reported negative outcomes for BMT. For instance, Patel [5] found that BMT implementation does not enable academics to achieve their desired teaching goals. Furthermore, Harkin [6] found that BMT implementation not only makes it difficult to interpret student and tutor behaviors but also to anticipate new demands and pressures placed on students and tutors. Due to the current gap in the literature, this study aims to investigate the impacts of BMT implementation on universities through multiple case studies. The study will explore various aspects such as students' performance, level of satisfaction, teacher support, and challenges that staff may encounter. Additionally, the study aims to provide insights for universities interested in adopting this teaching approach and for learners considering enrolling in BMT universities. The study question underlying this paper is:

RQ: What are the impacts of BMT projects on universities after implementation?

There is a gap in the literature regarding studies that offer a relatively comprehensive taxonomy of both positive and negative consequences that BMT can bring to universities. Therefore, by conducting this study, the authors aim to propose a comprehensive taxonomy for both the positive and negative impacts of implementing BMT in universities.

The rest of the paper is as follows: after this introductory section, a literature review will be presented, followed by a discussion of study methodology. Finally, the analysis and study findings will be presented, followed by a discussion of the findings and concluding remarks.

2 LITERATURE REVIEW

According to Davies [7], the BMT refers to subjects offered over a shorter period but more frequently and with longer hours. Although BMT has been implemented in several universities, the debate about what BMT truly offers has always been a concern. According to Tatum [3], students who attended BMT classes tend to have better learning outcomes. However, there are some limitations to these findings, such as memory retention, long-term outcomes, and course withdrawals. Since BMT courses are often brief, it is challenging for universities to gather sufficient information to compare with the traditional teaching mode. However, Burton & Nesbit [8] suggest that students who have chosen BMT or who have become familiar with BMT over time tend to perform better compared to students who have been studying in traditional classes.

Although there is very limited information regarding the implementation of BMT at universities, it is essential to learn from available study publications to gain a better understanding of BMT implementation projects and their impacts at various educational levels.

The study of the consequences of implementing BMT is quite limited. Among these studies, Karjanto and Yong [4] conducted a study at the University of Nottingham Malaysia Campus regarding the pros and cons of BMT for delivering mathematics modules to engineering students. They found that although BMT provides students and teachers with the opportunity to build their relationship and allows teachers to monitor students' progress closely, some students find it challenging to adapt. Lu and Wu [9] investigated an integrated evaluation approach to teaching and learning. The two researchers compared the traditional teaching mode with the new teaching approach that combines teaching and learning activities in BMT, enhancing learning activities and opportunities to interact with teachers. The results showed that students can acquire knowledge more effectively in the latter approach, where lessons are learned through open discussions, interactions with teachers, and numerous practices in the class.

Some other researchers collected data regarding the BMT implementation at Victoria University in the first-year college [10]. After analysing the data, they found that BMT implementation has positive impacts on students' learning outcomes. Some of these impacts include an increase in grades and the level of student engagement. Following the study of [10] on the BMT implementation project at Victoria University in the first-year college, Semerawickrema and Cleary [11] continued the study to evaluate the impacts of the project in the second year. The studies discovered that pass rates increased significantly, especially among students from non-English-speaking countries and those with low socio-economic status.

[12] investigated the impacts of BMT implementation on science subjects at the tertiary level. The study findings suggested that universities should only implement BMT on a subject-by-subject basis, as the block model is not suitable for all subjects. Patel [5] conducted a study on the post-implementation of BMT at Manchester Metropolitan University, where lecturers were invited to contribute their opinions regarding the impacts of the project. The study found that BMT implementation cannot provide academics with the opportunity to fulfill their desired teaching goals. Harkin [6], a lecturer in the Department of Psychology at Manchester Metropolitan University, investigated the impacts of the online block teaching model using Lefebvre's Trialectic of Space. Despite the fact that BMT implementation seems to allow universities to operate safely, it is not only difficult to interpret student and tutor behaviors but also challenging to foresee new demands and pressures that students and tutors are placed under.

Lostroh [13] provided more insights regarding BMT implementation at Colorado College. Their study found that although there are some concerns that block courses (usually taught in 3.5 weeks) cannot be delivered substantively, there are two strategies that can have positive impacts on students' deep learning. These strategies involve focusing on one area or delving deep conceptually and taking students off-campus to gain more practical knowledge. Additionally, peer-to-peer interactions are crucial to enhancing the effectiveness of BMT. However, there are still debates about whether BMT courses truly facilitate real learning or merely create an illusion. The table below summarizes the previous studies.

Table 1 summarizes what past studies have found on the topic of BMT consequences.

Table 1. Overview of the relevant research studies

Author (Name, Year)	Key Findings
Karjanto & Yong (2018) [4]	Provided the pros and cons of block teaching model for delivering mathematics modules to engineering students.
Lu & Wu (2018) [9]	Compared the traditional teaching mode and the new teaching approach where there is a combination of teaching and learning activities in BMT
McCluskey Weldon & Smallridge (2019) [10]	BMT implementation at Victoria University in the first-year college.
Semerawickrema & Cleary (2021) [11]	BMT implementation at Victoria University in the second-year.
Harvey, Power & Wilson (2017) [12]	The impacts of BMT implementation were studied on science subjects in tertiary level.
Patel (2021) [5]	A post-implementation study of BMT at Manchester Metropolitan University.
Harkin (2021) [6]	Investigated the impacts of online block teaching model by using Lefebvre's Trialectic of Space.
Lostroh (2007) [13]	Studied BMT implementation at Colorado Collage.

3 METHODOLOGY

This study employs an interpretivism research paradigm. This research paradigm posits the belief that study questions can be answered based on human experience and social reality. It utilizes a qualitative research approach. According to [14], a qualitative research approach is a qualitative inquiry that uses data as instruments to interpret and uncover underlying or hidden meanings. Humans are frequently at the core of qualitative research activities because qualitative research is commonly conducted to observe, interview, and analyze specific factors influencing human behaviors.

3.1 Research method and data collection

The study method applied in this study is a multiple-case study, which is an appropriate study approach to generate an in-depth, contextual, multi-faceted understanding of a complex real-world subject or issue. A case study enables the researcher to explore the key meanings, implications, and characteristics of the case by focusing on one case in depth or multiple cases to make comparisons in various aspects of the study question [15]. Different studies have defined case studies differently. For example, Stake [16] believes "A case study is both the process of learning about the case and the product of our learning" (p. 237). Miles and Huberman [17] describe it as "a phenomenon of some sort occurring in a bounded context" (p. 25), while Green and Thorogood [18] define it as an "In-depth study undertaken of one particular 'case', which could be a site, individual, or policy" (p. 284). However, the main concept is to conduct an in-depth investigation of an event or phenomenon and its impacts on the surrounding environment. Table 2 shows the results of the last step of data collection and the cases that have been targeted and studied.

As mentioned, this study employed the case study research method by studying five universities and higher education institutes that have already adopted BMT. These cases include Salford University, Colorado College, Heriot-Watt University, Victoria University, and Quest University. The data collected was secondary data from various verified, reliable, and valid sources such as university websites, higher education magazines, newspapers, non-profit media, university student blogs, academic publications, educational forums, higher education conferences, educational discussion groups, and so on. Table 2 briefly introduces these five cases. Each of the five cases is explained after Table 2.

Table 2. Cases studied in this study

Case Name	Country	The Highlight of the Case
Victoria University	Australia	VU has completely restructured its teaching style to BMT and the students need to complete subjects in 'blocks' of 4 weeks for undergraduate and 8 weeks for postgraduate. The undergraduate model involves studying one unit at a time, where the postgraduate involves studying 2 units at a time.
Salford University	UK	The school offers block teaching and learning by six weeks intensive programme which focuses on one subject only. There are also exceptions.
Colorado College	US	Students on the Block Plan take their subjects in three and a half weeks blocks, followed by a 4-day block break.
Heriot-Watt University	UK	They offer block teaching which is spread over a period of 6 weeks or less.
Quest University	Canada	They offer blocks of one month for each subject and it usually begins on a Monday and ends on the Wednesday of the fourth week, each class is of three hours duration.

Victoria University (VU), an Australian university, has completely restructured its teaching style by switching all its units to the BMT. Undergraduate and first-year college students are required to complete subjects in 'blocks' of four weeks, while postgraduate students have eight-week blocks. Results indicate that the block model has already delivered significant improvements for the university. VU has reported an overall increase in pass rate following the implementation of BMT. A notable impact on various student groups, such as those with non-English-speaking backgrounds, Indigenous students, and those from lower socio-economic backgrounds, has been reported by Loton [19].

Salford University applied BMT in a modular framework for various programs such as business and nursing. Each module is a self-contained block of learning and is designated at different levels: Level 3 (foundation level), Level 4 (certificate level), Level 5 (diploma level), Level 6 (degree level), and Level 7 (postgraduate level). In each block, students study one subject at a time in a six-week intensive program. They are assessed by two assignments, one after week 3 and one after week 6.

In Colorado College, students on the Block Plan take their subjects in three-and-a-half-week blocks, followed by a four-day block break. Currently, Colorado College offers 11 blocks. During the COVID-19 pandemic, the block model allowed the college to adopt a phased approach to bringing students back to campus. Colorado College is finding the ability to adapt and react rapidly as one of the benefits of adopting the block model of teaching.

In the BMT model adopted by Heriot-Watt University, classes are spread over a period of six weeks or less. The subject design is structured to provide students with 150 hours of learning for each subject. This model at Heriot-Watt University incorporates problem-based learning (PBL) sessions for each subject, giving students the opportunity to reflect on their learning.

At Quest University, a Canadian institution, students enrol in month-long blocks for each subject. These blocks typically start on Monday and conclude on Wednesday of the fourth week, with each class lasting three hours. Unlike traditional universities, at Quest University, students focus on one subject at a time. This unique approach allows students to collaborate around the clock and engage in self-directed projects within the framework of discipline-specific and interdisciplinary blocks.

3.2 Data analysis

This study utilized the thematic analysis method to analyze the data. Braun and Clarke [20] define thematic analysis as a method for identifying, analyzing, and reporting patterns within the collected data to extract relevant information. Various studies have employed different steps for thematic analysis. Consistent with prior studies [20], [21], and [20], this study followed the four key steps outlined below for data analysis:

Step 1: The initial step involves reading and re-reading the gathered data multiple times to help students become acquainted with the dataset. Once researchers are familiar with the collected data, they can acquire a more profound understanding of the BMT and fully comprehend the data.

Step 2: The second step involves coding or identifying the codes according to features or the information that the data carries.

Step 3: Step three involves examining the collected data along with the generated codes. Researchers should dedicate time to reviewing data codes repeatedly and consider whether certain codes may offer information related to a specific aspect of the project or fall under the same category. This particular aspect can be identified as a theme or pattern of information.

Step 4: The fourth step is to evaluate and outline the themes. A theme is a pattern that captures something important about the data or study question.

4 RESULTS

This section demonstrates all the findings of this study. The findings are responses to the study question and help understand the impacts of BMT implementation in the tertiary sector. Figure 1 presents the taxonomy, which has been created as a result of analyzing data collected through thematic analysis. This taxonomy shows the 21 impacts of BMT adoption (including 14 positive impacts and 7 negative impacts) grouped into eight categories. The impacts in yellow font in Figure 1 show the negative impacts, and those in white font are the positive impacts. Following Figure 1, each category and the underlying impacts in each category are explained.

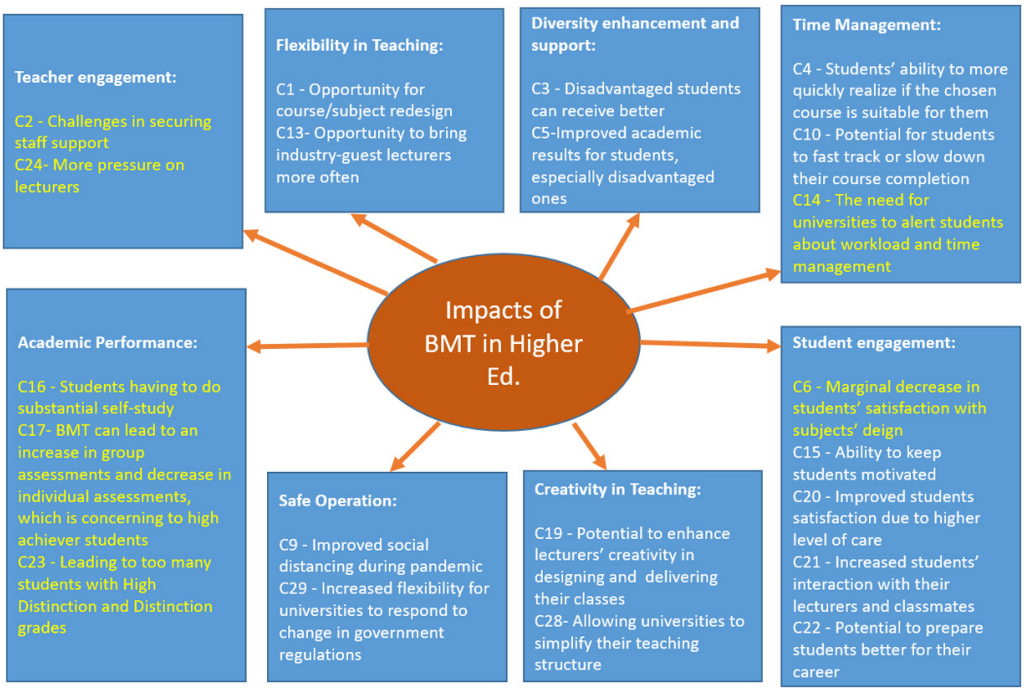

Fig. 1. Taxonomy of findings

4.1 Flexibility in teaching

The category 'flexibility in teaching' refers to any impacts that result in an increase or decrease in universities' flexibility in delivering their classes. This category has two underlying positive impacts of the Block model of teaching.

C1 – Opportunity for course/subject redesign. The BMT has the potential to provide universities with the opportunity to redesign courses and subjects to align with students' demands, ultimately leading to greater student retention. According to Ross (2018), universities are increasingly interested in modifying the way they deliver education to better suit students' preferred learning methods. Given that students may have diverse backgrounds and require additional support in learning, or may prefer shorter courses to accommodate work responsibilities, adapting classes to fit their needs appears to be a strategy for educational institutions to attract more student enrolment.

C13 – Opportunity to bring industry-guest lecturers more often. The BMT implementation projects allow universities to bring guest lecturers more frequently, including alumni who can share their work experiences and lessons with current students. According to [22], Salford Business School has implemented a program called the MBA Guest Leadership Speaker Series, which they have considered a module in the block class. In this module, students have the opportunity to interact with guest speakers, who are also Salford Business School alumni, and these speakers can share their knowledge and work experiences.

4.2 Teacher engagement

This category is used to describe any impacts the BMT implementation has on university staff. There are two negative impacts associated with this category.

C2 – Challenges in securing staff support. Block model of teaching has presented challenges for universities in obtaining staff support, as it is difficult to persuade staff to adopt and commit to a new teaching approach. An interview conducted by [23], titled 'Block teaching model ripe for copycats', highlighted that convincing employees to embrace and test the new teaching model after its implementation is challenging. Consequently, universities continue to encounter difficulties in securing support after implementing it, even after the Block model of teaching.

C24 – More pressure on lecturers. The BMT can lead to lecturers feeling more pressured as they have to finish all the markings and submit final results in a shorter timeframe. In a discussion under Dodd's article titled "Block teaching comes of age at Victoria University" [24], some people believed the amount of pressure the BMT projects put on teachers was tremendous. This is because teachers are required to not only redesign the classes but also take on longer classes and finish all the marking within a shorter period compared to traditional teaching methods. As the amount of work increases significantly, many teachers are forced to simplify the lesson content and assessments, which could be the reason why students' academic results increase significantly.

4.3 Diversity enhancement and support

This category pertains to the impacts of BMT implementation on disadvantaged students. It encompasses two positive impacts of the Block model of teaching.

C3 – Disadvantaged students can receive better support. The BMT has the potential to help students who have disadvantages in learning ability, are low performers, or have modest financial backgrounds, as they have a better chance of getting their questions answered in class.

The implementation of BMT at universities has had many positive impacts on students, especially first-in-family students, who are defined by O'Shea et al. [25] as the first persons in their immediate families of origin to attend an institution at the tertiary level. In an article written by Ross [23], Professor Dawkins states that the implementation of BMT projects brings many positive impacts on first-in-family students, who also represent a large group of students at Victoria Universities. As these students often face disadvantages in language, economics, or physical abilities, BMT allows the university to redesign the subjects to meet their needs, provide more opportunities for them to have their questions answered in class, and offer additional assistance and time to help them understand the learning concepts in class.

C5 – Improved academic results for students, especially disadvantaged ones. The BMT can generally lead to improved academic results for students who have economic or learning disadvantages. According to [26], students with a business background seem to enjoy greater positive impacts from the projects, while students from the arts and education disciplines seem to be less impacted. Although the reasons are not identified, they may be due to the nature of their disciplines or depending on their departments. Nevertheless, the overall level of satisfaction among this group of students increases.

4.4 Time management

The theme is the staff's ability to manage their time effectively in BMT. This category has three underlying impacts: two positive impacts and one negative.

C4 – Students' ability to more quickly realize if the chosen course is suitable for them. The BMT allows students to determine if their chosen area of study is suitable for them in a shorter period of time compared to the conventional model. This enables students to switch to another study area without wasting excessive time and money. When universities implement the BMT, it helps students ascertain if the chosen discipline is a good fit for them in a shorter time frame, as they dedicate a significant amount of time to learning and practicing in classes [23]. Consequently, students can transition to a more suitable discipline, saving time and money before it becomes too late.

C10 – Potential for students to fast track or slow down their course completion. The BMT has the potential to empower students who wish to take full control of the pace at which they complete their studies and how they manage their work-life balance.

According to [22], students studying in BMT can choose to complete their course faster or more slowly, depending on how they want to manage their life events or personal circumstances. While full-time students usually take two subjects per block, part-time students can take one subject and work part-time. In some situations, when students have life events that require their full attendance, they can simply skip one block and return to school the following block.

C14 – The need for universities to alert students about workload and time management. With BMT implementation, universities need to alert students about the workload required to complete assessments on time. According to James Cook University Australia [27], universities adopting BMT must warn students about the work required and meet assessment deadlines.

4.5 Student engagement

This category is used to describe the impacts of BMT implementation on student engagement. It includes five underlying impacts, four positives, and one negative.

C6 – Marginal decrease in students' satisfaction with subjects deign. While students' academic performance can generally improve by studying through BTM, they may experience a slight decrease in satisfaction with the design of subjects and a slight increase in satisfaction with their lecturers. According to [26], the impact of projects on students' performance across different disciplines is unclear, but students are believed to be slightly more content with their lecturers. However, their attitudes towards the design of their block units differ, as they do not provide much positive feedback regarding the design of these units.

C15 – Ability to keep students motivated. The BTM can help keep students motivated and interested in learning. According to Roseman University [28], students feel more motivated studying in BMT, as it allows them to focus on one unit at a time and understand the content completely before moving to a higher level of knowledge. In addition, all students are required to participate fully and contribute to class discussions. Therefore, their level of interest is also significantly increased while their knowledge is deepened.

C20 – Improved students' satisfaction due to higher level of care. The BTM can lead to higher satisfaction among certain groups of students due to the increased level of care and support provided to them. According to Curtin University [29], the implementation of BMT helped the university focus on their segmented group of students. Curtin University has a program called the Centre for Aboriginal Studies (CAS) that caters to students from Aboriginal communities. The implementation of BMT projects has enabled them to provide extra care to their students by redesigning the subjects to meet these students' needs and schedules.

C21 – Increased students' interaction with their lecturers and classmates. As a result of the BMT implementation, the size of classes has been reduced, allowing students to have more opportunities to interact with their lecturers and classmates. According to [30], BMT requires universities to reduce their class size as classes will be held more regularly, for longer hours, but within a shorter period. As students spend more time with each other and with their teachers, they have more opportunities to interact and discuss, which can help them build long-term relationships. This benefits students when they enter the workforce, as they already have connections in the field.

C22 – Potential to prepare students better for their career. The BTM has the potential to better prepare students for their careers. According to [30], BMT is believed to better prepare students for the job market as the length of the learning course is reduced.

4.6 Creativity in teaching

This category discusses how the implementation of BMT has influenced universities' creativity in delivering classes. There are two positive impacts associated with this category-block model of teaching.

C28 – Allowing universities to simplify their teaching structure. The BTM implementation projects have enabled universities to simplify their teaching structure, fostering creativity in learning and embracing students' potential rather than adhering to traditional academic measurement methods. According to Laville [31], after implementation, BMT has empowered teachers to embrace students' potential in learning by adopting more creative teaching approaches. Teachers can introduce various changes to the learning content within the required development principles, enabling them to emphasize students and devise more innovative methods to promote student learning. Consequently, new teaching approaches are being explored, leading to a significant enhancement in students' creativity.

C19 – Potential to enhance lecturers' creativity in designing and delivering their classes. The BTM has the potential to enhance lecturers' creativity in designing and delivering their classes and exploring different ways to improve student learning. According to Laville [31], after implementing BMT, teachers are impressed with the opportunities to be more creative and motivated, particularly in taking proactive steps to explore more creative teaching approaches.

4.7 Safe operation

This category is used to describe any impacts that BMT implementation has on universities' ability to operate safely and in accordance with government regulations. This category encompasses two positive impacts of the block model of teaching.

C9 – Improved social distancing during pandemic. The BTM has provided universities with the opportunity to continue operating safely during the pandemic by enabling them to eliminate large classes and reduce interaction between students. As noted by Nadworny [32], this is possible because class sizes can be reduced, teachers can opt to move classes online, or half of the learning is conducted online while the other half takes place on campus. Since students typically focus on one subject at a time, the ability to reduce interaction and trace COVID cases has been proven effective.

C29 – Increased flexibility for universities to respond to change in government regulations. The BTM can give universities more flexibility in responding to changes in government regulations, especially during pandemics, enabling them to continue their operations safely. As indicated by Manchester Metropolitan University [33] and the university teaching academy department [34], BMT allows universities to offer students "one unit at a time" to plan and act more swiftly.

4.8 Academic performance

This category is used to describe any impacts that BMT has on students' academic performance. This category encompasses three negative impacts.

C16 – Students having to do substantial amount of self-study. With BMT, students often must do a substantial amount of private study in their subjects to be able to fully understand everything. According to some nursing students who joined the forum discussion regarding the BMT on University Reviews [35], there are many flaws in the teaching materials that were simplified by their teachers as a result of BMT delivery. These students believed the knowledge they were taught in some blocked classes was not complete and not fully accurate and that they had to do a significant amount of self-directed learning.

C17 – BMT can lead to an increase in group assessments and decrease in individual assessments, which is concerning to high achiever students. The BMT can lead to an increase in the number of group assignments and a decrease in the number of individual assignments, which can, in turn, lead to an increase in high-achieving students' concerns as their grades now depend on other students who may not be learning at the same level as them. In the discussion forum under university review [35], when transitioning to BMT, teachers are compelled to simplify the learning content, reduce individual assignments, and increase the number of group projects. Consequently, some high-achieving students may struggle when assigned to a group where all the group members are not performing at the same level as them. As a result, these students tend to feel frustrated when they must either do more group work for the team or accept lower grades.

C23 – Leading to too many students with high distinction and distinction grades. The BMT can result in an increase in the number of students achieving high distinction grades. According to Marc, in the discussion group of the university review [35], students enrolled in BMT tend to achieve higher scores. The transition to BMT has led to a significant rise in the number of students passing the block course with high distinction. This can be attributed to simplified assessments and the introduction of more quizzes when adopting the block model of teaching.

5 DISCUSSIONS AND RESEARCH CONTRIBUTION

This paper aimed to uncover the positive and negative impacts of implementing BMT in the tertiary sector. Educational studies can benefit from this study by learning from its findings. They can gain more knowledge about the impacts that BMT implementation can have on universities. Studies can then consider our findings, compare them to their own study findings, and conduct further studies to explore different aspects of BMT implementation. Table 3 illustrates this study's contribution by showing which impacts confirm previous studies' findings and which impacts are new to the best of our knowledge.

Table 3. Research contribution

Impact Number	Newly Found (NF) OR Confirming Existing Literature
1	NF
2	NF
3	[36], [37], [38]
4	NF
5	NF
6	NF
9	NF
10	[39]
13	NF
14	NF
15	[40], [37, 41]
16	NF
17	NF
19	NF
20	[37]
21	NF
22	NF
23	[37]
24	NF
28	NF
29	NF

This study can bring enormous benefits to various groups of practitioners, including universities or educational providers, governments, students who are preparing to enter universities and colleges, and their parents. The study can benefit three categories of educational providers at different levels: universities or colleges that have already adopted BMT, those that are considering implementing BMT, and those that do not have a plan to adopt BMT but are interested in learning about it as an alternative.

As for the first group, universities and colleges that have already adopted BMT, this paper provides them with a relatively comprehensive taxonomy of the various positive and negative consequences of BMT implementation. It potentially communicates with them, makes them aware of some possible challenges that implementing BMT can bring about, and allows them to work on them. Since this study paper has collected a wide range of data, for example, from many discussion forums where students and staff are more willing to express their opinions, the findings from this paper can potentially provide these universities with insights they have not already been aware of. The taxonomy proposed herein also allows such institutes to examine their BMT implementation against the positive and negative impacts found to see how they have fared with the block model of teaching.

As for the second group, universities and colleges that already have plans to implement BMT, this paper provides them with an overview of the consequences they can expect from implementing BMT in their environment. This will help them leverage their resources to attain maximum benefits and minimize potential negative consequences.

As for the third group, universities and colleges that do not have a plan to adopt BMT but are interested in learning about BMT as an alternative, the taxonomy provided here can offer insights into the potential positive and negative outcomes that adopting BMT can bring. This will help them identify areas for improvement and anticipate challenges they may face if they choose to adopt the block model of teaching.

The taxonomy provided here can also inform government departments of education about policy-making for the tertiary sector. By considering our findings, governments can review their rules and regulations to assist universities in adopting teaching methods that best suit students.

Students who are about to enter colleges or universities, as well as their parents, are another group that can benefit from these study findings. By examining the impacts of studying in BMT on current students, these individuals can assess whether block courses align with their learning preferences before making a decision about which universities to attend. Parents can also utilize our findings to gain insights and engage in discussions with their children about the different universities they can select from.

6 CONCLUDING REMARKS

Block mode of teaching is a condensed teaching approach that enables universities to streamline the learning process and appeal to students who prefer this style of learning. BMT has shown significant potential for enhancing students' academic performance. This study utilized a multiple case study method, examining five higher education institutions from four different countries that have integrated BMT: Salford University, Colorado College, Heriot-Watt University, Victoria University, and Quest University. The thematic analysis of the gathered data resulted in the development of a taxonomy outlining 14 positive impacts and seven negative impacts of BMT implementation, categorized into 8 groups. The findings are valuable for various stakeholders, particularly universities, as they complement the adoption of the block model of teaching.

7 REFERENCES

[1] G. Cawelti, "High school restructuring: What are the critical elements?" *NASSP Bulletin,* vol. 79, no. 569, pp. 1–15, 1995. https://doi.org/10.1177/019263659507956902

[2] T. Murray, I. Barkat, K. Pearlman, and L. Robinson, "Intensive mode screen production: An Australian case study in designing university learning and teaching to mirror 'real-world' creative production processes," *Media Practice and Education,* vol. 21, no. 1, pp. 18–31, 2020. https://doi.org/10.1080/25741136.2019.1644842

[3] B. C. Tatum, "Accelerated education: Learning on the fast track," *Journal of Research in Innovative Teaching,* vol. 3, no. 1, pp. 35–51, 2010.

[4] N. Karjanto and S. T. Yong, "Some pros and cons of implementing parallel and block teachings for mathematics modules," *arXiv preprint arXiv:1810.07365*, 2018.

[5] H. Patel, "Decolonising the curriculum within a block teaching structure: A beginning," *Learning and Teaching in Action [Special issue],* vol. 14, no. 1, pp. 9–17, 2021.
[6] B. Harkin, "Academic readiness and collaborative communities in online block teaching: A Lefebvrian case study," *Learning and Teaching in Action*, vol. 14, pp. 1–8, 2021.
[7] W. M. Davies, "Intensive teaching formats: A review," *Issues in Educational Research,* vol. 16, no. 1, pp. 1–20, 2006.
[8] S. Burton and P. L. Nesbit, "Block or traditional? An analysis of student choice of teaching format," *Journal of Management & Organization*, vol. 14, no. 1, pp. 4–19, 2008. https://doi.org/10.5172/jmo.2008.14.1.4
[9] Y.-L. Lu and C.-W. Wu, "An integrated evaluation model of teaching and learning," *Journal of University Teaching & Learning Practice,* vol. 15, no. 3, p. 8, 2018. https://doi.org/10.53761/1.15.3.8
[10] T. McCluskey, J. Weldon, and A. Smallridge, "Re-building the first year experience, one block at a time," *Student Success,* vol. 10, no. 1, pp. 1–15, 2019. https://doi.org/10.5204/ssj.v10i1.1148
[11] G. C. Semerawickrema, "Block mode study: Opportunities and challenges for a new generation of learners in an Australian University," *Student Success,* vol. 12, no. 1, pp. 13–23, 2021. https://doi.org/10.5204/ssj.1579
[12] M. Harvey, M. Power, and M. Wilson, "A review of intensive mode of delivery and science subjects in Australian Universities," *Journal of Biological Education,* vol. 51, no. 3, pp. 315–325, 2017. https://doi.org/10.1080/00219266.2016.1217912
[13] C. P. Lostroh, "Keeping it real: Substantive learning on a short calendar," *CBE—Life Sciences Education,* vol. 6, pp. 285–288, 2007. https://doi.org/10.1187/cbe.07-05-0024
[14] S. B. Merriam and E. J. Tisdell, *Qualitative research: A guide to design and implementation*. John Wiley & Sons, 2015.
[15] S. Crowe and K. Sarma, "Coping with Covid-19: Stress, control and coping among pregnant women in Ireland during the Covid-19 pandemic," *BMC Pregnancy and Childbirth,* vol. 22, p. 274, 2022. https://doi.org/10.1186/s12884-022-04579-1
[16] R. E. Stake, *The art of case study research*. Sage, 1995.
[17] M. B. Miles and A. M. Huberman, *Qualitative data analysis: An expanded sourcebook*. Sage, 1994.
[18] J. Green and N. Thorogood, "Qualitative methods for health research," 2018.
[19] D. Loton, I. Solomonides, and M. Trish, "Building on the foundations: Results from the inaugural block cohort," in *HERDSA Conference 2019*, Auckland, New Zealand, 2019.
[20] V. Braun and V. Clarke, "Using thematic analysis in psychology," *Qualitative Research in Psychology,* vol. 3, no. 2, pp. 77–101, 2006. https://doi.org/10.1191/1478088706qp063oa
[21] M. Maguire and B. Delahunt, "Doing a thematic analysis: A practical, step-by-step guide for learning and teaching scholars," *All Ireland Journal of Higher Education,* vol. 9, no. 3, 2017.
[22] Moogan, "Why study for an MBA at Salford?" https://blogs.salford.ac.uk/business-school/why-study-for-an-mba/#more-9284
[23] J. Ross, "Block teaching model ripe for copycats," *Times Higher Education,* 2018. [Online]. Available: https://www.timeshighereducation.com/news/block-teaching-model-ripe-copycats
[24] T. Dodd, "Block teaching comes of age at Victoria University," 2020.
[25] S. O'Shea, J. May, C. Stone, and J. Delahunty, "First-in-family students, university experience and family life: Motivations, transitions and participation," *Palgrave Macmillan*, 2024. https://doi.org/10.1007/978-3-031-34451-0
[26] J. Ross, "Block teaching elevates marks by a grade, study finds," *Times Higher Education,* 2020. [Online]. Available: https://www.timeshighereducation.com/news/block-teaching-elevates-marks-agrade-studyfinds

[27] James Cook University, "Intensive mode subjects," 2021. [Online]. Available: https://www.jcu.edu.au/students/enrolment/block-mode-subjects
[28] Roseman University, "The differences between block curriculum and traditional curriculum delivery," 2014. [Online]. Available: https://www.roseman.edu/2014/11/02/the-differences-between-block-curriculum-and-traditional-curriculum-delivery/
[29] O. V. Curtin University, "Block release intensive study format," 2021.
[30] T. Dodd, "Two more universities are adopting block teaching," *The Australian,* 2020.
[31] H. Laville, "See what can be done," *Learning and Teaching in Action,* vol. 14, no. 1, 2020.
[32] E. Nadworny, "6 Ways college might look different in the fall," *NPR,* 2020. [Online]. Available: https://www.npr.org/2020/05/05/848033805/6-ways-college-might-look-different-in-the-fall
[33] Manchester Metropolian University n.d., "Learning and teaching in action," *University Teaching Academy,* 2021. [Online]. Available: https://www.celt.mmu.ac.uk/ltia/BTcall.php
[34] U. T. A. n.d., "What is block teaching?" *University Teaching Academy,* 2021. [Online]. Available: https://www.uta.mmu.ac.uk/block_teaching/resources/what-is-block-teaching/index.html#/
[35] U. R. n.d., "University reviews," Victoria University. https://universityreviews.com.au/students/vic/victoria/
[36] R. Klein, K. Kelly, P. Sinnayah, and M. Winchester, "The VU way: The effect of intensive block mode teaching on repeating students," *International Journal of Innovation in Science and Mathematics Education,* vol. 27, no. 9, 2019. https://doi.org/10.30722/IJISME.27.09.004
[37] C. W. Yao and C. Collins, "Perspectives from graduate students on effective teaching methods: A case study from a Vietnamese Transnational university," *Journal of Further and Higher Education,* vol. 43, no. 7, pp. 959–974, 2019. https://doi.org/10.1080/0309877X.2018.1429583
[38] M. Swain, "Block teaching and the three A's: Attendance, attainment and attitudes," *Innovations in Practice,* vol. 10, no. 1, pp. 33–38, 2016.
[39] D. B. Grant, "Using block courses for teaching logistics," *International Journal of Physical Distribution & Logistics Management,* vol. 31, nos. 7–8, pp. 574–585, 2001. https://doi.org/10.1108/09600030110402987
[40] S. Bourgeois, N. Drayton, and A.-M. Brown, "An innovative model of supportive clinical teaching and learning for undergraduate nursing students: The cluster model," *Nurse Education in Practice,* vol. 11, no. 2, pp. 114–118, 2011. https://doi.org/10.1016/j.nepr.2010.11.005
[41] A. K. Kofinas, Y. Bentley, and C. Minett-Smith, "Block teaching as the basis for an innovative redesign of the PG suite of programmes in university of Bedfordshire Business School," 2017.

8 AUTHORS

Nguyen Thi Thao Trinh is with the College of Sport, Health and Engineering, Victoria University, Melbourne, Australia (E-mail: thi.trinh24@live.vu.edu.au).

Amir Hossein Ghapanchi is with the College of Sport, Health and Engineering, Victoria University, Melbourne, Australia; The Institute for Sustainable Industries and Liveable Cities, Victoria University, Melbourne, Australia (E-mail: Amir.Ghapanchi@vu.edu.au).

Afrooz Purarjomandlangrudi is with the College of Arts, Business, Law, Education and IT, Victoria University, Melbourne, Australia (E-mail: afrooz.purarjomandlangrudi@vu.edu.au).

Online-Journals.org

iJET International Journal of Emerging Technologies in Learning

iJET | eISSN: 1863-0383 | Vol. 19 No. 6 (2024) | OPEN ACCESS

https://doi.org/10.3991/ijet.v19i06.50165

PAPER

Where Is the Teacher in Data Analytics in Education? Evaluating the Maturity of Analytics Solutions and Frameworks Supporting Teachers

Ngoc Buu Cat Nguyen[1](✉), **Thashmee Karunaratne**[2]

[1]University West, Trollhättan, Sweden

[2]Royal Institute of Technology (KTH), Stockholm, Sweden

ngoc.buu-cat-nguyen@hv.se

ABSTRACT

COVID-19 has changed the mindset of many teachers from traditional education to online education. The increased use of learning management systems is leveraging opportunities for increased use of learner data to draw insights about the learners and the learning environment. However, typically learners are the primary beneficiaries, while teachers are quite invisible in the research of data analytics in education, although both are equally important. Thus, this paper aims to position teachers in the spotlight by differentiating between these current two definitions of learning analytics (LA) and teaching analytics (TA) and evaluating the applicability and maturity of existing analytics solutions to support teachers in making decisions on teaching and learning. A systematic literature review was conducted in relevant scientific fields. The results showed clear evidence to distinguish TA from LA and that there are only a few TA solutions and frameworks that can be applied widely or in reality. Evaluating TA solutions and frameworks needs to be attentively considered. This paper also contributes a comprehensive TA process framework that encapsulates the missing elements in the previous models and adds the recent highlights raised in the fields. The implications for research and practice are also discussed.

KEYWORDS

teaching analytics, learning analytics, learning design, teachers, design research cycle

1 INTRODUCTION

During COVID-19, many teachers had to quickly transfer to online education [1]. This led to the growth of data creation by students and teachers on learning management systems, opening more chances for learning and teaching enhancement through data analytics. The tendency toward online or blended education and increased use of education technology systems in teaching and learning continues

Nguyen, N.B.C., Karunaratne, T. (2024). Where Is the Teacher in Data Analytics in Education? Evaluating the Maturity of Analytics Solutions and Frameworks Supporting Teachers. *International Journal of Emerging Technologies in Learning (iJET)*, 19(6), pp. 19–37. https://doi.org/10.3991/ijet.v19i06.50165

Article submitted 2024-05-17. Revision uploaded 2024-05-21. Final acceptance 2024-05-21.

after the pandemic, which increases the advantages of drawing implicit insights from data. This paper contributes to the theoretical perspective of supporting teachers in making decisions on teaching and learning through data analytics by differentiating between the current two popular terms in the fields: learning analytics (LA) and teaching analytics (TA), which is still a debate in the respective research fields. Concurrently, this paper also provides an evaluation of the applicability and maturity of the existing TA solutions and frameworks for supporting teachers and teaching.

LA emerged in tandem with the surge of data in education and with a focus on students and their learning experiences. This aligns well with the definition of LA by the Society for Learning Analytics Research "LA is the measurement, collection, analysis, and reporting of data about learners and their contexts for purposes of understanding and optimizing learning and the environments in which it occurs" [2]. Accordingly, the questions related to teachers and teaching are not directly included or addressed in the definition. The discourse of data-informed decision-making to support the quality of learning and teaching through pedagogical innovations, timely feedback, or enhanced learning design (LD) can be found in contemporary LA research. However, the substantial focus of this discourse turns toward students and addresses the questions of students rather than the dilemmas that teachers encounter. The language used in the LA field does not speak directly to the concerns and issues that teachers grapple with [3]. The results supporting teachers and teaching are not emphasized clearly and explicitly enough in the LA papers. Therefore, we argue that a direct and explicit focus on teachers is also of equal demand. Based on this fact, TA emerged to stress the issues and questions of teachers and teaching directly. However, the conceptualization of TA is still in its infancy [4] compared to the maturity of the LA research.

[4] defined TA as "a reconceptualization of LA for teachers to improve teaching performance and learning outcome." According to [5], TA focuses on "the design, development and evaluation of notations, representations, and visualizations of learning and teaching processes and products and the enculturation of a 'professional vision' for teachers to make the visual analytics notations, representations, and visualizations meaningful and actionable in pedagogical settings." Based on the definitions of LA and TA, the two key actors—students and teachers—are clearly stated and separated in each term. However, the related literature still describes the methodological approaches that have been implemented aiming at supporting teachers or improvements for LD as the ultimate goals under LA (ex. [6]). As a result, the controversy is still ongoing, which is, on one hand, that TA is viewed as a sub-field of LA since they see TA from the aspect of LA in connection with the teacher inquiry process [7]; on the other hand, LA and TA are considered as two separate research fields [4, 5].

In contemporary literature, there are two ways to find TA papers. In some publications, the term TA is explicitly used as a keyword. In other relevant publications, TA is implicit in the discourse of LA to improve LD. Additionally, a major challenge for LA and TA is to realize the promise of developing actionable interventions based on analytical results to improve learning and teaching, which is often alluded to as closing the loop [8]. Nevertheless, the loop has rarely been closed since many studies were apt to stay in the early stages [7, 9], resulting in a shortage of practical and successful evidence of TA and LA applications [8]. The first aim of this paper is to draw a clear line between the terms LA and TA and evaluate the maturity of TA solutions based on closing the loop, leading to the first research question:

RQ1: What are the differences between LA and TA? How mature are the TA solutions to be applied in practice?

In line with the discourse in RQ1, we also investigate the maturity of the frameworks for TA and LD supporting teachers (including teaching and environments) when designing courses integrated with data analytics due to the fact that it is still a question of how generalizable and applicable the frameworks work in authentic settings and in widespread practice [10]. Hence, the second purpose of this paper is to synthesize and evaluate the existing frameworks for TA and LD to provide a holistic perception of what the research community offers to support teachers with LD, leading to the second research question:

RQ2: How generalizable and applicable are the existing frameworks for TA and LD to come into use?

2 RESEARCH METHOD

We conducted a systematic literature review (SLR) as a methodology in the databases, covering the relevant, well-known outlets for identifying the differences between LA and TA and the applicability of the existing frameworks for TA and LD.

2.1 Article search strategy

[4] conducted an SLR to contextualize the notion of TA and develop various concepts around TA. This study, however, has not investigated the maturity of the TA solutions and their relevance for enhancing teaching practices in general and LD in particular. Additionally, [9] did SLR research to evaluate the development status of LA software concerning LA-driven LD improvements in higher education by using the software development life cycle as a reference model. To move forward from these two papers which have been done in a similar topic or direction, we clarify the TA field in this paper by conducting an extensive SLR to distinguish TA from LA, delving into the analysis of the development process of TA solutions that the studies have gone through, and evaluating the applicability and maturity of existing TA solutions to support teachers to enhance LD in various contexts (not limited to higher education), as well as the existing frameworks for TA and LD. This approach has not been used in previous literature. This review intends to offer various stakeholders, especially teachers, learning designers, and those working with LD, a landscape of TA-integrated LD, how applicable TA solutions are to improve LD and be applied in practice, and the perception of what has been explored, implemented, and validated in TA-integrated LD.

In the identification of relevant articles, the databases covering the publication venues in the subjects, including EBSCO, ACM, IEEE Explore, Science Direct, SpringerLink, Scopus, and the Journal of LA, were examined. The search queries primarily contained two key topics: “teaching analytics” and “learning design”. Due to the immaturity of the TA field, the term “learning analytics” was also searched together with “learning design” since this combination often leads to papers benefiting teachers and teaching. Moreover, the other keyword “instructional design” was included since it was interchanged with “learning design” in some papers. The inclusion criteria and the details of the SLR are depicted in Table 1. To find the right and relevant articles, the search scope in the databases was selected within keywords, subjects, or abstracts, depending on the available option of the advanced

search function in each database. Two databases, SpringerLink and the Journal of LA, did not have the advanced search function, so the search scope could not be considered.

Table 1. Details of the systematic literature review

Databases	Search Queries	Inclusion Criteria	Search Scope	Results
EBSCO	("learning analytics" OR "teaching analytics") AND ("learning design" OR "instructional design")	Peer reviewed, articles or conference papers, in English, not duplicated, searched in keywords, between 2010 to 2022	Subjects	81
ACM Digital Library			Abstract	17
IEEE Explore			Keywords	16
ScienceDirect			Keywords	9
SpringerLink			Not available	779
Scopus			Keywords	163
Journal of Learning Analytics			Not available	19
			Total	1084

2.2 Coding scheme and eligibility criteria

The selected corpus of articles was analyzed using the document analysis method. The coding scheme consisted of 19 aspects to record the relevant methodological and substantial features of the studies in table form. These aspects include stakeholders, theories, research questions, country, education level, subject, contribution/outcomes, differences between TA and LA, dataset size, data, implementation, technique/algorithms, current issues in analytics and pedagogy, evidence of analytics in LD, reach intervention, implications, limitations, future work, and kind of artifact. However, only relevant aspects that helped answer the research questions were considered.

According to the search strategy described above, the total number of papers was 1084 from the databases with the inclusion criteria. These papers went through the PRISMA diagram to select the final set of papers. The first step was to remove duplications using the Rayyan website for all the databases except SpringerLink. 149 duplicates were found. Regarding SpringerLink, this database did not support the function of filtering out incomplete papers such as lecture notes, in-progress work, workshop papers, and posters. Thus, the researcher examined the results from SpringerLink manually by skimming titles and abstracts to remove the incomplete papers. In the second step, the remaining results from SpringerLink and the other databases were combined and skimmed through titles and abstracts to eliminate papers irrelevant to teachers, LD, and LA or TA. In the third step, the 189 remaining articles were full-text read, coded, and analyzed. Among the 189 articles, the researcher continued to exclude the irrelevant papers that were not found by skimming titles and abstracts. In the set of irrelevant papers, although the papers regarding learning spaces, teacher professional developments, or tool evaluation aimed at teachers, they were not counted due to beyond the research focus of this paper. Some papers explored teacher education, which was pertinent to LA, TA, and LD, yet the key target was students in teacher education, so these papers were disregarded. As a result, the articles using LA/TA to improve LD, implementing teacher-facing dashboards or tools for teachers to create LDs, or developing

a framework of analytics and LD were taken into account. After all, 104 hits were relevant for the in-depth reading and analysis of the content. In the last step, the researcher included 21 papers with TA solutions which came to the stage of evaluation or intervention and 11 papers with evaluated frameworks for LA/TA and LD to present in the result section since these papers help answer the RQs. To clarify why the researcher divided into two sets of final results for two RQs, for RQ1, the researcher considered the studies using TA methods to generate solutions for the identified problems, thereby measuring the maturity of TA solutions. RQ2 stressed the frameworks for TA/LA and LD; accordingly, it was not necessary to use TA methods to develop the frameworks, but the frameworks can be developed based on theories, previous literature, or qualitative methods. In spite of this reason, there were five common papers in both final sets of articles. As explained, for each RQ, the papers were analyzed following different aspects; thus, there was no conflict for these common papers. Figure 1 summarizes the bibliographic information for the selection process.

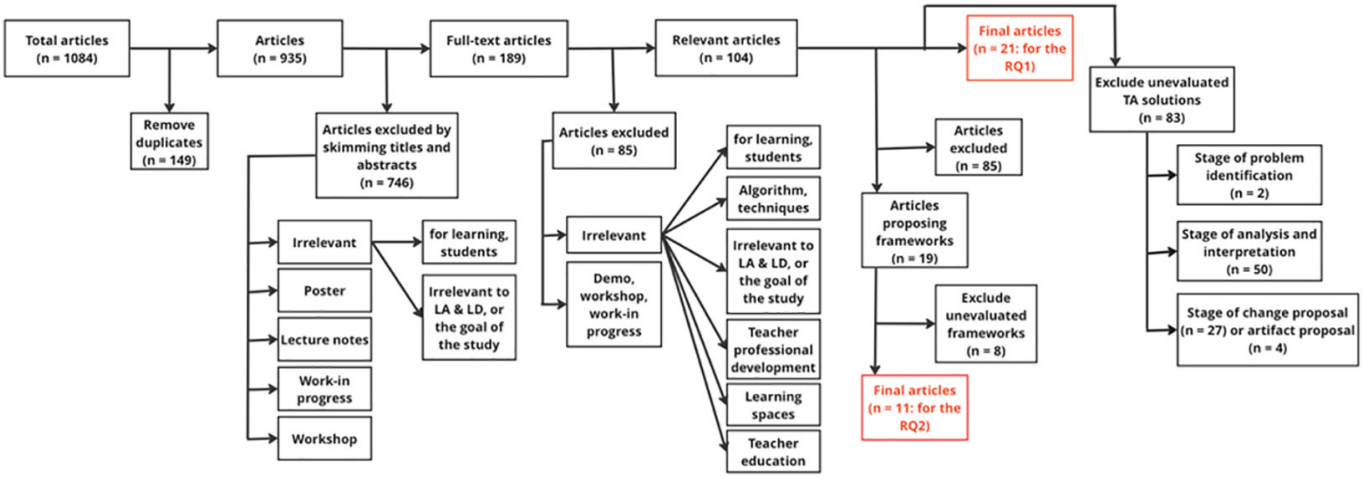

Fig. 1. PRISMA flow diagram of the screening and selection

2.3 Selecting a reference model for data analysis

To choose a method for data analysis, the researcher looked into the set of 104 relevant articles to explore a proper reference model as the baseline to measure the maturity and applicability of the TA solutions presented in the selected literature. The most common and specific methods that ten papers followed to develop an artifact or improve LD are design science research, design-based research, action research, or similar processes that were not explicitly named as the design cycle. The other methods are relatively generic and imply other various approaches included (more details in Table 2). Thus, we chose the design research cycle for the analysis of this SLR to investigate the differences between LA and TA based on this cycle and evaluate the applicability and maturity of TA solutions. For the papers that did not explicitly use the design cycle, we examined their studies based on the steps they had taken to discover how far their research had gone and how applicable the solutions were outside their own case studies. Some authors combined several methods in their studies (e.g. [11, 12]), leading to a total of more than 104 papers in Table 2.

Table 2. Methods with a respective number of papers

Qualitative method	7	Mixed method	8	Design science research/ design-based research/ action research	10
Case study	8	Review and theory	12	Flipped classroom	1
Multiple steps	9	Experiment	2	Phenomenological approach	1
SLR	9	Multimodal LA	2	Multiple studies	1
Non-specific method	36	Multimethod study	1		

To analyze the set of 21 papers for RQ1, we developed a taxonomy of elements following the design process (refer to Table 3). We did not apply this taxonomy to the set of 11 papers with frameworks for the RQ2 because we analyzed the frameworks in terms of applicability and generalizability. Regarding the taxonomy, one cycle means that authors systematically conceptualized, designed, developed, and evaluated the solutions once. Two cycles represent the solutions that are evaluated once and revised, or that led to interventions for improving the learning-related functionality that the articles address. The design research cycle, as a reference model, includes specific steps of awareness of problems, suggestion, development, evaluation, and conclusion [13]. Due to the distinction of this design cycle, only the papers producing artifacts were appropriate and able to be measured based on this design process. TA solutions and frameworks are artifacts in this context and embodied in different ways, such as insights, principles, tools, or prototypes. The contemporary research studies were analyzed to discover how many papers have reached the evaluation/intervention stage or how many cycles the papers went through, which shows the maturity of the models as described above.

The first element of the taxonomy is foundation, referring to whether a study is formed based on either theory, literature, or problems from authentic contexts, or even both. The second element, the design research cycle, signifies how many cycles of the design research cycle a study has gone through to evaluate the maturity of the development process of the study. The element of evaluation extent elucidates to what extent an artifact is tested and evaluated. The last element, applicability, alludes to the capacity that an artifact can be applied in reality, to a small or large extent.

Table 3. A taxonomy of elements following design research cycle

	Low	Medium-Low	Medium	High
Foundation	Non foundation	Problem or tendency with literature	Problem with potential solutions and strong literature/theory	Theory and empirical evidence
Design research cycle	1 incomplete cycle	1 complete cycle	2 cycles	>2 cycles
Evaluation extent	Test based on literatures or low-fidelity prototypes	Partly test or ongoing test in reality	Test with one course or limited group in reality	Test with more than one course or in wide scope
Applicability	Unsure applicability in authentic settings or do not know yet	Potential but uncertain or limited in reality	Considerable extent of applicability and readiness in reality with limitations or unexpected effects	Widely applied, high reliability, or high applicability with minor issues

2.4 An overview of the maturity of TA solutions

The set of 104 relevant articles has been read to capture an overview of the maturity of TA solutions. Among them, two papers have come to the stage of understanding the problem and context [14] or investigating design principles but have not reached the implementation stage [15]. Almost half of the review list—50 articles—only reported the outcomes up to the level of interpreting analytics results aligning to context and LDs, or proposed an artifact, even a handcrafted artifact or an incomplete artifact. These papers did not discuss intervention or evaluation to measure the applicability of the findings. Thus, these research studies did not show evidence for a complete design cycle, resulting in unanswered questions about the applicability of the outcomes or the reproducibility of the methods in other learning settings than the example showcased in the respective study.

Additionally, 27 papers proposed changes for LDs after analyzing and interpreting the results while four papers proposed an artifact with the illustration of use cases to show how the artifacts work. These papers have not come to the evaluation stage of the artifacts in reality or the intervention to apply proposed changes. The paper of [16] also needed further clarification although the authors used the term "intervention" in the paper. Referring to the design-based research approach that the authors used in the paper, the authors identified learning barriers and proposed revisions for the course design strategies in the stage of analysis and exploration. Next, the authors implemented the changes for the course design strategies, which were well-adopted in the stages of design and construction. However, the system construction and evaluation were in planning as future work. Thus, this paper was classified in the group that proposed changes. Another paper doing an SLR summarized the current landscape of LA and LD [17]. Although this paper did not produce any artifacts, the authors suggested several important areas of investigation to move forward from the current state. Thus, this paper was put in the category of proposing changes and used as the ground for our study.

Interestingly, 21 papers have reached the stage of evaluation or intervention. These papers are presented in detail in the result section. Figure 2 provides a comprehensive insight into classifying these 104 articles.

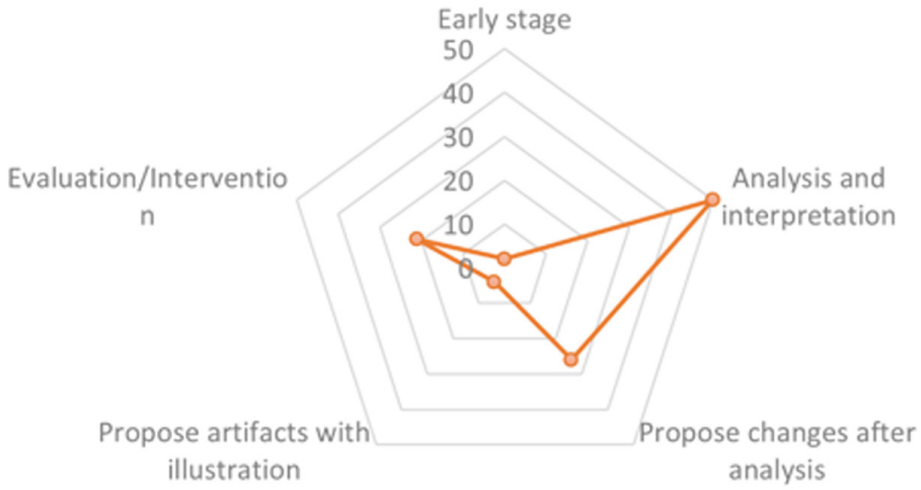

Fig. 2. A chart of articles classified by the design research cycle

2.5 An overview of existing frameworks for TA/LA and LD

As described in Figure 1, there were 19 papers proposing frameworks or models of TA/LA-integrated LD for teachers. These frameworks are listed in Figure 3.

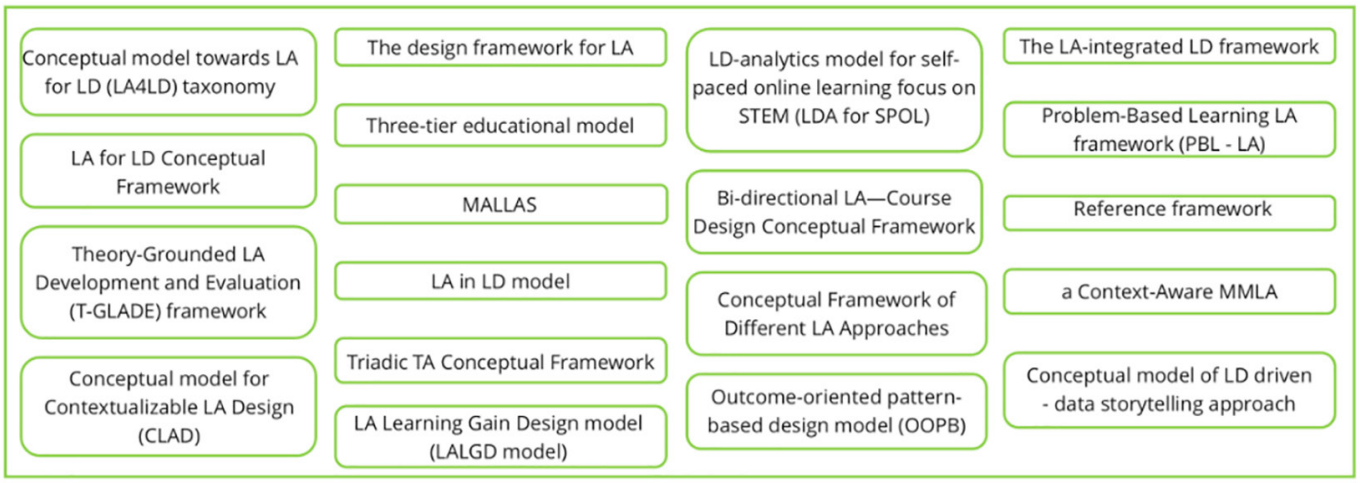

Fig. 3. Existing TA/LA and LD frameworks/models

Six papers proposed frameworks based on literature, which have not been evaluated. One paper evaluated their framework through the Loop tool, which was in progress [18]. [19] put forward a reference framework and used it as the analytical framework for the SLR of the same paper. The evaluation of this framework was not explicitly shown but implied potential. These eight frameworks have not been evaluated, so they were not analyzed. The remaining eleven frameworks have been evaluated either using use cases, literature, or real-world cases, which are presented in the result section.

3 RESULTS

3.1 RQ1: How mature are the TA solutions to be applied in practice?

The RQ1 has two parts. In this section, we respond to a part of the question about the maturity of TA solutions to be able to be applied in practice. 21 papers were found that have reached the stage of evaluation or intervention. The taxonomy details of these papers are summarized in Table 4. Among them, [20] brought up the analytical tool SNAPP which provides real-time visualizations of discussion activities. To examine the performance of this tool as reported in the literature, two more subsequent versions from these authors were found, which helped perceive the entire picture of the tool development process.

Among the articles in our corpus, three papers focus on designing learning objects or learning activities rather than the entire LD. Namely, [21] focused on the design of a writing activity with the support of an automated writing feedback tool; [22] worked on online asynchronous discussions with the support of Starburst; and [23] concentrated on revising content through the dashboard CoReaDa.

There is a sequence of papers that have connections in terms of the design process. The sequence starts with proposing design principles for edCrumble in [24],

followed by [25], and finally [26]. The papers were counted separately and presented in Table 4.

Table 4. Results of articles completing at least one design cycle

	Foundation	Design Process	Evaluation Extent	Effectiveness
Molina-Carmona et al. [27]	Medium-Low	High	Medium	Medium
Lockyer and Dawson [20]	Medium-Low	Not clear	High	Medium
Schmitz et al. [10]	Medium	High	High	High
Echeverria et al. [28]	Medium	Medium-Low	Low	Low
Blumenstein [29]	Medium	Medium-Low	Low	Low
Zotou et al. [12]	Medium	Medium-Low	High	Medium
Ahmad et al. [19]	Medium	Medium-Low	Medium	Medium-Low
Kaliisa and Dolonen [30]	Medium	High	Medium	Medium
Li et al. [31]	Medium	High	Medium	Medium
Eradze et al. [32]	High	Medium	Medium	Medium
de Menezes et al. [33]	Medium-Low	Medium	Medium	Medium
Kitto et al. [34]	Medium	Medium	Medium	Medium
Albó et al. [26]	High	Medium-Low	Medium	Medium
Hilliger et al. [35]	Medium	Medium	High	High
Albó and Hernández-Leo [24]	Medium-Low	Medium-Low	Medium-Low	Low
Albó et al. [25]	Medium	Medium-Low	Medium	Medium
Ortega-Arranz et al. [36]	High	Medium	Medium-Low	Medium-Low
Rodríguez-Triana et al. [37]	High	High	High	High
Sadallah et al. [23]	Medium-Low	Medium-Low	Medium	Medium
Shibani et al. [21]	Medium	Medium-Low	High	Medium
Dietrich et al. [38]	Medium	Medium-Low	High	Medium-Low

According to Table 4, three of 21 studies showed the high applicability and reliability of the outcomes, which were assessed based on their broad evaluations with various contexts, stakeholders, and their multi-iteration design processes. Among those three articles, [35] reported that the tool is being used by 20 Latin American universities to inform teachers about learning situations when redesigning course assessments and sequences. The other two still suggested future work to make the artifacts better at using TA to improve LD. Moreover, the SNAPP tool is also widely used as a reflective tool instead of for modification of the LD "on the fly" as the initial expectation [20]. The remaining papers need to make adjustments to enhance the alignment of TA and LD in their studies.

3.2 RQ2: How generalizable and applicable are the existing frameworks for TA and LD to come into use?

Eleven evaluated frameworks are presented here to answer RQ2. One framework was evaluated based on 13 selected studies from the SLR [29]. Three papers

demonstrated how to implement the frameworks through use cases [8, 39, 40]. One framework was offered based on the two-cycle design of an empirical study [6], another paper demonstrated three iterations to evaluate the frameworks [31], and five frameworks have been tested in reality from one course to multiple courses. Table 5 shows the details of the eleven frameworks evaluated in authentic settings or through use cases.

Table 5. The details of usability and applicability of the evaluated frameworks

Frameworks	LD	Subject	Learning Mode	Data Size
CLAD by Shibani et al. [21]	Writing activity	Various	Blended, online	90 and 302 university students
LDA for SPOL by Yan et al. [16]	Self-pace	Computer science	Online	University students
PBL-LA by Zotou et al. [12]	Problem-based learning	Various	Classroom with e-learning environment	32 postgraduate students
Context-aware MMLA by Eradze et al. [32]	LD for secondary teaching	Not specific	Classroom	1200 secondary students
LD driven-data storytelling approach by Echeverria et al. [28]	Group work	Database	Classroom with e-learning environment	15 undergraduate students
OOPB by Li et al. [31]	Not specific	E-learning leadership	Blended	21 master students
Design framework for LA by Seufert et al. [40]	**1.** Forum discussion **2.** Reflection **3.** The use of game elements **4.** With materials and problems	Not specific	Online	**1.** A group **2.** At individual level **3.** Community or individual learners **4.** Large number of learners
Multilevel framework for LA integrated LD by Law and Liang [8]	Task: observe the scenarios in the presented stimulus.	STEM	Classroom with e-learning environment	Individual student in grade 8
Bi-directional LA-course design by Kaliisa et al. [39]	Online social interaction activities and contexts	Not specific	Blended	University learners
T-GLADE by Wiley et al. [6]	Assessments: multiple choice items and open-response items	Global climate change	Classroom with e-learning environment	885 middle school students
LALGD model by Blumenstein [29]	Socio-communicative, cognitive/metacognitive, metacognitive/affective	Not specific	Face-to-face, blended, online	Not empirical evaluation but based on 13 selected studies, in higher education

According to Table 5, the LALGD model [29] was validated using literature but not in an authentic setting, so it is unclear to indicate for which subject or data size this model is valid, rather than the contexts applying the LDs of socio-communicative, cognitive/metacognitive, or metacognitive/affective. This model was created for higher education contexts; accordingly, the other educational contexts are not applicable. The frameworks described by [28, 39, 40] can be generalized to various LDs due to the popularity of these activities. However, the authors did not particularly describe how the activities should be designed and implemented. [40] put a heavier focus on LA design, while [39] demonstrated quite a general LD, and [28] emphasized the effects between explanatory and exploratory visualizations in the dashboard. The frameworks by [12, 16, 21] focused on specific kinds of LDs, which limits the generalizability of these frameworks. Context-aware MMLA [32] fits well in a physical setting rather than in

an online or blended context. The OOPB model [31] can be generalized since it can be applied to different educational levels, different modes, various subjects, and stresses on the activities that produce patterns. The authors of the OOPB model created four taxonomies of design patterns, including directed learning, explorative learning, productive learning, and reflective learning; thereby, instructional designers or teachers can freely refer to according to their design plans. Additionally, it is uncertain if the frameworks described by [6, 8, 32] function in higher education as they have been evaluated in secondary and middle schools. Five frameworks [8, 12, 28, 31, 40] can handle small datasets with a small number of students (<50), while the rest manage large datasets or do not mention this information explicitly.

3.3 TA or LA: which one supports teachers?

By synthesizing the knowledge captured from the SLR, we answer the remaining part of the RQ1 about the differences between LA and TA in sections 3.3 and 3.4. We start with the definitions of TA and LA introduced in the introduction, LA and TA are distinguished by two key actors: LA for learners and TA for teachers. We can examine two directions as follows.

In the first direction, the study of [21] used LA to build an automatic writing feedback tool tuned with pedagogical context to support personalized learning and to address students' dilemma: "How can students improve their writing skills without teachers?" The authors closely aligned the purpose of this study with the LA definition of aiming for students and their learning and emphasized contextualizable LA design to produce meaningful support for students.

In the second direction, the study of [27] also utilized LA to analyze students' data to support teachers in redesigning an instructional course and provide in-time assistance, thereby addressing teachers' question, which is "in which ways teachers can motivate students' learning, especially low-performance students." The other study by [41] used LA to analyze students' communication and collaboration in order to help teachers reflect on their practices and LDs, thereby anticipating problems and making informed interventions. This study also answered teachers' questions, which are "How can teachers actually design and evaluate their course design?" and "How can teachers provide in-time support to their students?"

Comparing these two directions, teachers' questions are different from students' questions although both types of studies used students' data. The goals of these studies aimed at different actors and answered different questions. Accordingly, it is recognized that the second direction is incongruous with the LA definition but consistent with the TA definition. [28] revealed the differences between students' and teachers' concerns and questions through student-facing dashboards and teacher-facing dashboards. LA dashboards allow students to evaluate some parts of their learning behaviors and may even assist them in better managing their study, while TA dashboards help teachers obtain a thorough grasp of their entire course or specific tasks, reflect on their teaching strategies, and identify students who need particular support. There are also disparities between learners' and designers' perspectives. Teachers or designers desire to communicate multiple insights or dimensions of data about student experience to make better-informed decisions on (re)designing learning activities while students want to see what they are supposed to do, how they perform individually and in comparison to their classmates, when activities are happening, and if there are alternative ways to achieve the same goal [14]. Moreover, there is also the opinion that LA was never intended to exclude teachers, and it is obvious since

teachers and students always need to go hand in hand to improve both learning and teaching. One cannot support students without considering teachers and LD, and vice versa. However, what we argue here is that the LA definition focuses on students and learning and addresses students' questions, while the TA definition actually answers teachers' pedagogical questions and directly supports teachers. The ultimate result of both TA and LA benefits students in a direct way for LA and in an indirect way for TA through supporting teachers. Therefore, it is important to demystify that if a study aims to support teachers using data analytics, TA should be used, and if a study aims to support learners using data analytics, LA should be used. The differences between LA and TA are also divulged when we dig into the analysis of developing TA solutions based on the design research cycle in the next section.

3.4 Differences between LA and TA – proposing a TA process framework

We continue to answer RQ1 about the differences between LA and TA by analyzing two more perspectives to distinguish LA and TA, which are recognized throughout the design research cycle.

The first aspect deals with the data. It is remarked that TA and LA both used student data to reflect on and evaluate either LD or students' learning or give feedback to teachers. Yet, this review explored further that TA uses not only student data but also the data of teachers using tools [25], LD data [42], physical traces of students or teachers [28], or qualitative data. TA can use LD data and teachers' data to reflect on teaching practices or evaluate LD, learning activities, or learning objects, while LA cannot use LD data or the data of teachers using tools to assess students' learning.

In the second aspect, if a design process just stops at the analysis and interpretation stages, TA and LA look the same. Nevertheless, the difference lies in the later stages of the design process when analytical results are transferred to changes or interventions. Regarding TA, only teachers themselves can implement the changes in their LD to observe the impacts while students cannot implement changes in LD or teaching practices, except that students can give feedback on LD. Similarly to LA, students themselves can change their behaviors or learning ways to improve their learning, which teachers cannot do, except that teachers can provide motivation or support. The later stages of the design cycle clearly show the differences: TA generates impacts on LD, or teaching, and directly involves teachers, while LA produces impacts on learning and directly involves students. This argument is supported by the 21 papers reaching the later stages reported above.

There are also differences between the TA process and the LA process. Currently, there is no well-established or well-defined TA process. [11] named a teacher inquiry cycle including problems and questions, design intervention and evaluation, classroom implementation, collected data and analysis, and reflection and changes. The first and last steps can go back and forth before starting a new cycle. Both teachers and students are engaged in the process. [41] described a teaching cycle as including design and planning, engaging with students, reflection, professional development, and then starting a new cycle. This cycle is valid for various levels of granularity, from learning activities to sessions, modules, or programs. [4] designed the teaching outcome model (TOM) – a TA life cycle, comprising data collection, data analysis, data visualization, and data action (course design and assessment). In this model, the teacher is the central actor. While both the teacher inquiry cycle and teaching cycle did not clearly show TA but demonstrated well the aspect of the teacher's role, TOM reflected the components of the data analytics process but lacked the teachers' practices. Thus, premised on these cycles and this SLR, we aggregated and supplemented to enhance

the TA process framework (Figure 4). Our TA process framework encapsulates the missing elements in the previous models described above by bringing the teacher role to the center of the didactics, teachers' actions into the TA process, and adding up-to-date needs relating to TA as well as relevant stakeholders in tandem.

The starting point of this TA cycle is that teachers capture interests in learning and teaching based on their LD and available data concerning LD, characteristics of learners, situations of learners, environments of learning and teaching, as well as challenges that learners or teachers face. Continued is extracting and analyzing the respective log data of teachers, learners, or LD data following the interests. The data collection and analysis need to be performed in (near) real-time to provide continuous insights to teachers throughout the courses. In addition, the choice of appropriate algorithms, techniques, or methods conforming to the data is essentially taken into account. TA results are visualized and interpreted by connecting back to the LD and context to be able to perform meaningful and timely interventions. Based on the interpretation of TA results, teachers alone or together with researchers and analysts propose appropriate changes that can enhance the LD for enhanced learning. Depending on the kinds of changes that will be applied to a course, teachers, together with relevant stakeholders such as a department, study administrators, learners, or institutions, will engage in planning and (re)designing the course. Stakeholders involved in this step need to consider learning objects in the LD in such a way that TA can be used to assess or validate the effectiveness of the learning objects. Additionally, the TA-integrated LD needs to be grounded in theory, as suggested in the previous literature. The next step is to implement the changes in authentic courses, followed by teachers' as well as the relevant stakeholders' reflections and evaluations on the changes. A new loop will be formed following teachers' needs for improvement. In this cycle, teachers, analysts, and researchers can be either different people or one person if the person can play the roles and have the competencies of teachers, analysts, and researchers at the same time.

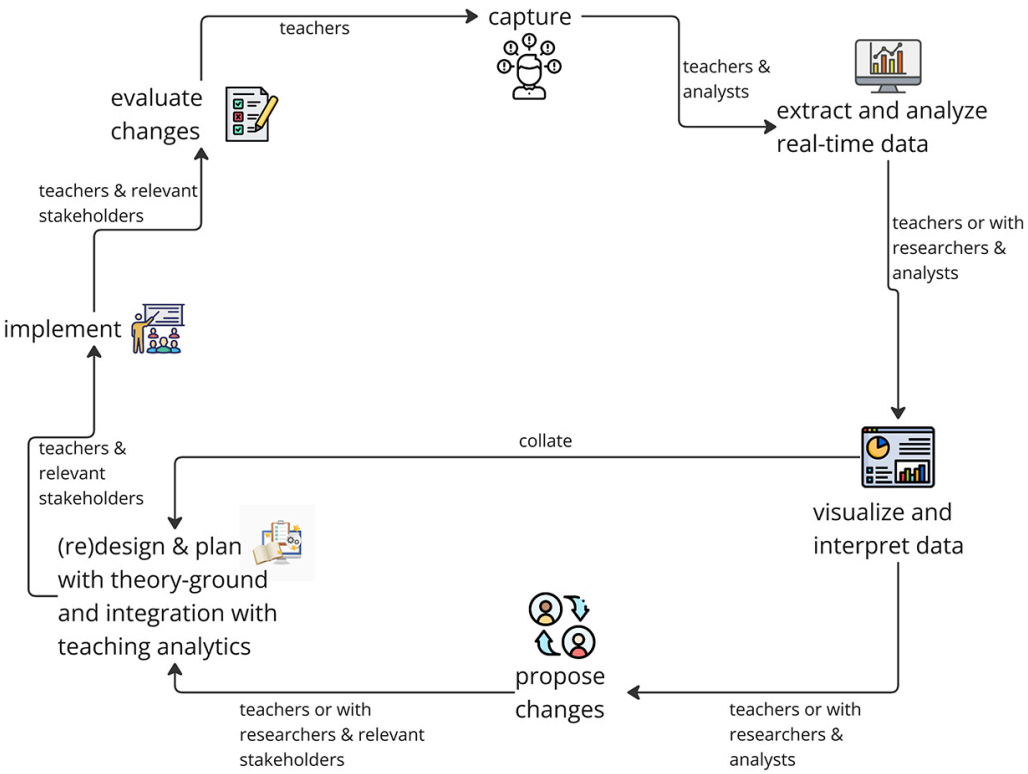

Fig. 4. TA process framework

4 DISCUSSION AND CONCLUSIONS

It is not uncommon to navigate students in the research of data analytics in education, but it is quite vague to see teachers as a key actor and their role. Although teachers still implicitly exist in this field, they should be emphasized and paid attention to as visibly and equally as students. Thus, we contribute this paper toward positioning teachers in this research field by presenting the results of an SLR aimed at demystifying the differences between TA and LA and delineating the applicability and maturity of TA solutions for teachers to improve LD and of the existing frameworks of TA-integrated LD. We selected and analyzed 21 articles developing TA artifacts to reinforce and supplement the points distinguishing between TA and LA in relation to LD. The design research cycle was applied for the data analysis as a reference model to examine the maturity and applicability of TA artifacts. This SLR showed that most of the existing efforts in TA focus on the first steps of this cycle, namely, data gathering, analysis, and visualization of certain learning processes rather than intervention or evaluation, which were illuminated in [7, 9]. However, some studies reached the intervention and evaluation steps, and three artifacts demonstrated high effectiveness or wide applicability in practice.

To differentiate between LA and TA, this paper revealed the ways, including (1) premised on the design research cycle, (2) kinds of data, and (3) definition, central actor, and actor's perspectives. According to the design research cycle, TA and LA are clearly distinguished in the later stages of the design process thanks to intervention while both TA and LA are quite similar in the early stages. Regarding the kinds of data, TA utilizes not only student data but also teacher data and data relating to teaching practices while LA works mostly with student data. Concerning the definitions, TA prioritizes teacher orientation whereas LA focuses primarily on learner orientation. The perspectives and questions of teachers and learners are different.

Moreover, this paper also synthesized and supplemented the elements to improve the TA process framework to be comprehensive and correlate with the recent highlights in the disciplines of TA and LD. Hopefully, this TA process framework combining teachers' practices and TA as well as highlighting teachers' roles can urge the studies using TA and aiming at teachers and LD to move forward closing the loop. Future suggestion is to evaluate the proposed TA process framework in real-world cases.

4.1 Implications for practice

Real-time feedback enhances the success of closing the loop [8]. Currently, a scarcity of real-time analysis hinders teachers from making quick decisions and accessing critical information. Based on the lessons learned from the case of the SNAPP tool [20] and the finding on the limited use of dashboards by teachers [30], to operate the real-time analysis and increase the possibility for adoption, this requires researchers' attention and the cooperation of teachers in designing TA-integrated activities and using tools to make it happen. Teachers' professional development in new pedagogies and data literacy should be prioritized to remove the barriers preventing teachers from using TA solutions and enable necessary intervention [43].

4.2 Implications for research

To develop TA solutions and frameworks, the design research cycle is a typical method that several papers have used compared to the other methods. The design research cycle is a rigorous process using theoretical knowledge and engineering design principles to construct socio-technical artifacts to solve identified problems. Evaluation is emphasized as crucial in this cycle [44]. Accordingly, the design research cycle is appropriate to use as not only an approach for developing artifacts but also a reference model to analyze the maturity of TA artifacts. The variety of artifact types and evaluation methods in this cycle facilitates the wide coverage of measuring the development and evaluation of TA solutions and frameworks.

There are existing frameworks combining TA and LD, as shown above. Nevertheless, the challenge is that many of them have not been evaluated; thus, the validity and reliability of these frameworks when applied in practice cannot be guaranteed. It is observed that most of the frameworks were implemented as tools. As a result, evaluating the frameworks connotes evaluating the respective tools; thereby, the evaluation of the frameworks is postponed until the technical implementation of the tools is accomplished [18]. Some of these frameworks were designed for specific learning activities such as writing or problem-based learning while others were tested in middle and secondary schools. This prevents these frameworks from being generalized; accordingly, these frameworks need to be evaluated in more contexts. Most of the existing frameworks support online learning or e-learning environments, which are comparatively popular in higher education. However, many courses in higher education have a small number of learners and, consequently, little data for building TA solutions. This is not supported or taken into account by most of the current frameworks. Conversely, there are potential frameworks that can be considered in the aspect of generalizability such as the OOPB model, bi-directional LA-course design, design framework for LA, or LD-driven data storytelling approach, yet they need more examination. Due to the limited contexts of evaluation, the maturity of these frameworks is not high for ubiquitous application. Consequently, there is an urgent need to explore the flexibility of LD for supporting diverse learners and teachers [17]. We also agree with [8, 10] that these frameworks are relatively complex and would not be ready to be adopted by practitioners. Hence, future work should consider the optimization of potential frameworks toward simplicity, generalizability, and applicability to support teachers, especially non-technical teachers.

5 REFERENCES

[1] S. Mehtälä, M. Salo, A. Rinne, T. Koskelainen, and H. Pirkkalainen, "Schoolteachers' experiences of the challenges associated with information technology use in traditional and online education," in *14th Scandinavian Conference on Information Systems*, 2023, vol. 11. https://aisel.aisnet.org/scis2023/11

[2] SoLAR, "What is learning analytics?" Society for Learning Analytics Research (SoLAR). https://www.solaresearch.org/about/what-is-learning-analytics/ (Accessed: 10/07, 2022).

[3] N. Law, L. Li, L. F. Herrera, A. Chan, and T.-C. Pong, "A pattern language based learning design studio for an analytics informed inter-professional design community," *Interaction Design and Architecture(s),* no. 33, pp. 92–112, 2017. https://doi.org/10.55612/s-5002-033-005

[4] I. G. Ndukwe and B. K. Daniel, "Teaching analytics, value and tools for teacher data literacy: A systematic and tripartite approach," *International Journal of Educational Technology in Higher Education,* vol. 17, no. 22, 2020. https://doi.org/10.1186/s41239-020-00201-6

[5] R. K. Vatrapu, "Towards semiology of teaching analytics," in *Workshop Towards Theory and Practice of Teaching Analytics, at the European Conference on Technology Enhanced Learning, TAPTA*, 2012, vol. 12.

[6] K. J. Wiley, Y. Dimitriadis, A. Bradford, and M. C. Linn, "From theory to action: Developing and evaluating learning analytics for learning design," in *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge*, 2020, pp. 569–578. https://doi.org/10.1145/3375462.3375540

[7] L. P. Prieto, K. Sharma, P. Dillenbourg, and M. Jesús, "Teaching analytics: Towards automatic extraction of orchestration graphs using wearable sensors," in *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge*, 2016, pp. 148–157. https://doi.org/10.1145/2883851.2883927

[8] N. Law and L. Liang, "A multilevel framework and method for learning analytics integrated learning design," *Journal of Learning Analytics,* vol. 7, no. 3, pp. 98–117, 2020. https://doi.org/10.18608/jla.2020.73.8

[9] E. Drugova, I. Zhuravleva, U. Zakharova, and A. Latipov, "Learning analytics driven improvements in learning design in higher education: A systematic literature review," *Journal of Computer Assisted Learning,* vol. 40, no. 2, pp. 510–524, 2023. https://doi.org/10.1111/jcal.12894

[10] M. Schmitz, M. Scheffel, R. Bemelmans, and H. Drachsler, "FoLA 2—A method for co-creating learning analytics-supported learning design," *Journal of Learning Analytics,* vol. 9, no. 2, pp. 265–281, 2022.

[11] I. Amarasinghe, K. Michos, F. Crespi, and D. Hernández-Leo, "Learning analytics support to teachers' design and orchestrating tasks," *Journal of Computer Assisted Learning,* pp. 1–16, 2022. https://doi.org/10.1111/jcal.12711

[12] M. Zotou, E. Tambouris, and K. Tarabanis, "Data-driven problem-based learning: Enhancing problem-based learning with learning analytics," *Educational Technology Research and Development,* vol. 68, pp. 3393–3424, 2020. https://doi.org/10.1007/s11423-020-09828-8

[13] B. Kuechler and V. Vaishnavi, "On theory development in design science research: Anatomy of a research project," *European Journal of Information Systems,* vol. 17, no. 5, pp. 489–504, 2008. https://doi.org/10.1057/ejis.2008.40

[14] M. Schmitz, M. Scheffel, E. van Limbeek, R. Bemelmans, and H. Drachsler, "'Make it personal!'-Gathering input from stakeholders for a learning analytics-supported learning design tool," in *Lifelong Technology-Enhanced Learning, EC-TEL 2018,* in Lecture Notes in Computer Science, V. Pammer-Schindler, M. Pérez-Sanagustín, H. Drachsler, R. Elferink, and M. Scheffel, Eds., Springer, Cham, vol. 11082, 2018, pp. 297–310. https://doi.org/10.1007/978-3-319-98572-5_23

[15] A. J. Edson and E. D. Phillips, "Connecting a teacher dashboard to a student digital collaborative environment: Supporting teacher enactment of problem-based mathematics curriculum," *ZDM–Mathematics Education,* vol. 53, pp. 1285–1298, 2021. https://doi.org/10.1007/s11858-021-01310-w

[16] H. Yan, F. Lin, and Kinshuk, "Including learning analytics in the loop of self-paced online course learning design," *International Journal of Artificial Intelligence in Education,* vol. 31, pp. 878–895, 2021. https://doi.org/10.1007/s40593-020-00225-z

[17] L. P. Macfadyen, L. Lockyer, and B. Rienties, "Learning design and learning analytics: Snapshot 2020," *Journal of Learning Analytics,* vol. 7, no. 3, pp. 6–12, 2020. https://doi.org/10.18608/jla.2020.73.2

[18] A. Bakharia *et al.*, "A conceptual framework linking learning design with learning analytics," in *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge*, 2016, pp. 329–338. https://doi.org/10.1145/2883851.2883944

[19] A. Ahmad *et al.*, "Connecting the dots–A literature review on learning analytics indicators from a learning design perspective," *Journal of Computer Assisted Learning,* 2022. https://doi.org/10.1111/jcal.12716

[20] L. Lockyer and S. Dawson, "Learning designs and learning analytics," in *Proceedings of the 1st International Conference on Learning Analytics and Knowledge*, 2011, pp. 153–156. https://doi.org/10.1145/2090116.2090140

[21] A. Shibani, S. Knight, and S. B. Shum, "Contextualizable learning analytics design: A generic model and writing analytics evaluations," in *Proceedings of the 9th International Conference on Learning Analytics & Knowledge*, 2019, pp. 210–219. https://doi.org/10.1145/3303772.3303785

[22] F. Marbouti and A. F. Wise, "Starburst: A new graphical interface to support purposeful attention to others' posts in online discussions," *Educational Technology Research and Development,* vol. 64, pp. 87–113, 2016. https://doi.org/10.1007/s11423-015-9400-y

[23] M. Sadallah, B. Encelle, A.-E. Maredj, and Y. Prié, "Towards fine-grained reading dashboards for online course revision," *Educational Technology Research and Development,* vol. 68, pp. 3165–3186, 2020. https://doi.org/10.1007/s11423-020-09814-0

[24] L. Albó and D. Hernández-Leo, "Identifying design lrinciples for learning design tools: The case of edCrumble," in *Lifelong Technology-Enhanced Learning: 13th European Conference on Technology Enhanced Learning, EC-TEL 2018,* Leeds, UK, September 3–5, 2018, Proceedings 13, Springer, 2018, pp. 406–411. https://doi.org/10.1007/978-3-319-98572-5_31

[25] L. Albó, J. Barria-Pineda, P. Brusilovsky, and D. Hernández-Leo, "Concept-level design analytics for blended courses," in *Lifelong Technology-Enhanced Learning, EC-TEL 2018*, in Lecture Notes in Computer Science, V. Pammer-Schindler, M. Pérez-Sanagustín, H. Drachsler, R. Elferink, and M. Scheffel, Eds., Springer, Cham, vol. 11082, 2018, pp. 541–554. https://doi.org/10.1007/978-3-030-29736-7_40

[26] L. Albó, J. Barria-Pineda, P. Brusilovsky, and D. Hernández-Leo, "Knowledge-based design analytics for authoring courses with smart learning content," *International Journal of Artificial Intelligence in Education,* vol. 32, pp. 4–27, 2022. https://doi.org/10.1007/s40593-021-00253-3

[27] R. Molina-Carmona, C. Villagrá-Arnedo, F. Gallego-Durán, and F. Llorens-Largo, "Analytics-driven redesign of an instructional course," in *Proceedings of the 5th International Conference on Technological Ecosystems for Enhancing Multiculturality*, 2017, pp. 1–7. https://doi.org/10.1145/3144826.3145391

[28] V. Echeverria, R. Martinez-Maldonado, S. B. Shum, K. Chiluiza, R. Granda, and C. Conati, "Exploratory versus explanatory visual learning analytics: Driving teachers' attention through educational data storytelling," *Journal of Learning Analytics,* vol. 5, no. 3, pp. 73–97, 2018. https://doi.org/10.18608/jla.2018.53.6

[29] M. Blumenstein, "Synergies of learning analytics and learning design: A systematic review of student outcomes," *Journal of Learning Analytics,* vol. 7, no. 3, pp. 13–32, 2020. https://doi.org/10.18608/jla.2020.73.3

[30] R. Kaliisa and J. A. Dolonen, "CADA: A teacher-facing learning analytics dashboard to foster teachers' awareness of students' participation and discourse patterns in online discussions," *Technology, Knowledge and Learning,* vol. 28, pp. 937–958, 2023. https://doi.org/10.1007/s10758-022-09598-7

[31] L. Li, L. Farias Herrera, L. Liang, and N. Law, "An outcome-oriented pattern-based model to support teaching as a design science," *Instructional Science,* vol. 50, pp. 111–142, 2022. https://doi.org/10.1007/s11251-021-09563-4

[32] M. Eradze, M. Rodríguez-Triana, N. Milikic, M. Laanpere, and K. Tammets, "Contextualising learning analytics with classroom observations: A case study," *Interaction Design & Architecture (ID&A),* no. 44, pp. 71–95, 2020. https://doi.org/10.55612/s-5002-044-004

[33] D. A. T. de Menezes, D. L. Florêncio, R. E. D. Silva, I. D. Nunes, U. Schiel, and M. S. de Aquino, "DaVid—a model of data visualization for the instructional design," in *2017 IEEE 17th International Conference on Advanced Learning Technologies (ICALT)*, Timisoara, Romania, 2017, pp. 281–285. https://doi.org/10.1109/ICALT.2017.97

[34] K. Kitto, M. Lupton, K. Davis, and Z. Waters, "Incorporating student-facing learning analytics into pedagogical practice," in *ASCILITE 2016, Conference Proceedings Show Me The Learning,* S. Barker, S. Dawson, A. Pardo, and C. Colvin, Eds., Adelaide, 2016, pp. 338–347. https://doi.org/10.14742/apubs.2016.810

[35] I. Hilliger, C. Aguirre, C. Miranda, S. Celis, and M. Pérez-Sanagustín, "Lessons learned from designing a curriculum analytics tool for improving student learning and program quality," *Journal of Computing in Higher Education,* vol. 34, pp. 633–657, 2022. https://doi.org/10.1007/s12528-022-09315-4

[36] A. Ortega-Arranz, P. Topali, J. I. Asensio-Pérez, S. L. Villagrá-Sobrino, A. Martínez-Monés, and Y. Dimitriadis, "e-FeeD4Mi: Automating tailored LA-informed feedback in virtual learning environments," in *Educating for a New Future: Making Sense of Technology-Enhanced Learning Adoption (EC-TEL 2022),* in Lecture Notes in Computer Science, I. Hilliger, P. J. Muñoz-Merino, T. De Laet, A. Ortega-Arranz, and T. Farrell, Eds., Springer, Cham, vol. 13450, 2022, pp. 477–484. https://doi.org/10.1007/978-3-031-16290-9_39

[37] M. J. Rodríguez-Triana, A. Martínez-Monés, J. I. Asensio-Pérez, and Y. Dimitriadis, "Scripting and monitoring meet each other: Aligning learning analytics and learning design to support teachers in orchestrating CSCL situations," *British Journal of Educational Technology,* vol. 46, no. 2, pp. 330–343, 2015. https://doi.org/10.1111/bjet.12198

[38] J. Dietrich, F. Greiner, D. Weber-Liel, B. Berweger, N. Kämpfe, and B. Kracke, "Does an individualized learning design improve university student online learning? A randomized field experiment," *Computers in Human Behavior,* vol. 122, p. 106819, 2021. https://doi.org/10.1016/j.chb.2021.106819

[39] R. Kaliisa, A. I. Mørch, and A. Kluge, "'My point of departure for analytics is extreme skepticism': Implications derived from an investigation of university teachers' learning analytics perspectives and design practices," *Technology, Knowledge and Learning,* vol. 27, pp. 505–527, 2022. https://doi.org/10.1007/s10758-020-09488-w

[40] S. Seufert, C. Meier, M. Soellner, and R. Rietsche, "A pedagogical perspective on big data and learning analytics: A conceptual model for digital learning support," *Technology, Knowledge and Learning,* vol. 24, pp. 599–619, 2019. https://doi.org/10.1007/s10758-019-09399-5

[41] Q. Nguyen, B. Rienties, and L. Toetenel, "Unravelling the dynamics of instructional practice: A longitudinal study on learning design and VLE activities," in *Presented at the Proceedings of the Seventh International Learning Analytics & Knowledge Conference (LAK '17)*, 2017, pp. 168–177. https://doi.org/10.1145/3027385.3027409

[42] G. Pishtari, M. J. Rodríguez-Triana, L. P. Prieto, A. Ruiz-Calleja, and T. Väljataga, "What kind of learning designs do practitioners create for mobile learning? Lessons learnt from two in-the-wild case studies," *Journal of Computer Assisted Learning,* pp. 1–17, 2022. https://doi.org/10.1111/jcal.12672

[43] Y. Wen and Y. Song, "Learning analytics for collaborative language learning in classrooms," *Educational Technology & Society,* vol. 24, no. 1, pp. 1–15, 2021. https://www.jstor.org/stable/26977853

[44] K. Peffers, M. Rothenberger, T. Tuunanen, and R. Vaezi, "Design science research evaluation," in *Design Science Research in Information Systems. Advances in Theory and Practice (DESRIST 2012),* in Lecture Notes in Computer Science, K. Peffers, M. Rothenberger, and B. Kuechler, Eds., Springer, Berlin, Heidelberg, vol. 7286, 2012, pp. 398–410. https://doi.org/10.1007/978-3-642-29863-9_29

6 AUTHORS

Ngoc Buu Cat Nguyen is a PhD candidate in Informatics, University West, Sweden (E-mail: ngoc.buu-cat-nguyen@hv.se).

Thashmee Karunaratne is an associate professor in Digital Learning, KTH, Sweden.

Online-Journals.org

iJET International Journal of Emerging Technologies in Learning

iJET | eISSN: 1863-0383 | Vol. 19 No. 6 (2024) | OPEN ACCESS

https://doi.org/10.3991/ijet.v19i06.49829

PAPER

Integration of AI and Metaheuristics in Educational Software: A Hybrid Approach to Exercise Generation

Blanka Láng(✉), Balázs Dömsödi

Corvinus University of Budapest, Budapest, Hungary

blanka.lang@uni-corvinus.hu

ABSTRACT

This study explores the integration of generative artificial intelligence (AI) with Exercise Generation Algorithm+ (EGAL+), a multi-objective harmony search (HS) metaheuristic-based algorithm capable of composing high-quality exercises. These exercises are characterized by their diversity, consistent difficulty, and comprehensive coverage of the source material, tailored to user preferences. One of the main challenges of using metaheuristics to compile exercises efficiently is the initial creation of a large question bank, which often demands significant time and effort from instructors. To overcome this challenge, the integration of a readily available existing generative AI module is proposed. This module is accessed through its application programming interface, autonomously populating the question bank. This sets the stage for EGAL+ to fine-tune the selection and assembly of specific exams. The resulting program enables educators to create an extensive question bank from any educational material, independent of the subject, and subsequently compose exercises with minimal effort. This approach leverages the synergistic benefits of both generative AI and metaheuristic-based optimization, offering a robust and efficient solution for exercise generation.

KEYWORDS

automatic question generation (AQG), artificial intelligence (AI), multi-objective optimization, exercise generation, metaheuristics, harmony search (HS)

1 INTRODUCTION

The manual generation of high-quality test sets, whether single or multiple-choice question (MCQ) sets, is a tedious and time-consuming task. Consequently, a significant amount of research has been dedicated to automatic question generation (AQG) to save teachers considerable time and effort. Compiling question sets is currently a routine task for many educators, who then utilize them for assessments. However, creating high-quality question sets is a laborious process involving numerous repetitive tasks. As teachers' time and energy could be better utilized in tasks that leverage their expertise to create more value, many algorithmic solutions for question

Láng, B., Dömsödi, B. (2024). Integration of AI and Metaheuristics in Educational Software: A Hybrid Approach to Exercise Generation. *International Journal of Emerging Technologies in Learning (iJET)*, 19(6), pp. 38–51. https://doi.org/10.3991/ijet.v19i06.49829

Article submitted 2024-03-24. Revision uploaded 2024-05-10. Final acceptance 2024-05-10.

set generation have been developed to streamline this process [1], [2], and [3]. The issue became more pronounced during the COVID-19 pandemic, with a surge in online exams conducted without in-person supervision [4], [5].

Some studies present general question-generating solutions that can be applied across all fields, while others design question-generating solutions tailored to specific disciplines [5], [6], and [7]. There may be significant variations in effectiveness when these solutions are applied to different fields. Some researchers concentrate solely on question generation, while others also examine the answers derived from the questions. The following recent publications review the literature and explore future research directions in the field of question generation.

A methodology is presented that combines generative software engineering principles with the process of feature-oriented product line engineering, specifically for question generation, as detailed in [4]. This generator allows for the creation of families of single-choice questions based on written templates, defined features, and parameters. According to the authors, examiners and instructors can easily utilize this generator to generate various question variants.

A model for generating programming questions that utilizes the local knowledge graph and abstract syntax tree is proposed in [7]. The generated questions were well-received, earning positive feedback from a group of experienced instructors.

In [8], a practical toolkit for question-and-answer generation (QAG) is provided. Generating questions with answers from a text can be useful in many areas, but it is not easy to achieve. This paper introduces an online service for multilingual QAG for end-users and a comprehensive Python package for fine-tuning and generating models for developers.

Several research papers have addressed the challenges and opportunities of utilizing large language models (LLMs) [2], [5], [9], [10], and [11]. In [9], it is argued that this fundamental technology is a key driver of innovation and that its application in education can bring many benefits. The research explores how these models can help create educational content and engage learners. It is pointed out that the use of LLMs in education requires participants to understand both the technology and its limitations.

In [10], the issue of MCQs generated from textbooks is explored in the context of using LLMs. The study investigates whether LLMs, due to recent advancements, can produce questions that are comparable to those generated by humans. Two LLMs are analyzed, highlighting their distinct issues: Macaw tends to duplicate answer choices, while Bing Chat often omits the correct answer from the choices provided. The findings indicate that the performance of the two LLMs does not significantly differ from human performance.

In [5], the study explores the generation of MCQs using LLMs specifically for computer science courses. It highlights that crafting high-quality MCQs is a time-consuming task and that LLMs are efficient in assisting instructors in creating questions that align with courses and learning goals. The findings indicate that GPT-4 outperformed GPT-3 in generating answers solely based on the question text.

In [12], the study explores the utilization of AQG Web services and AQG models to produce a Quran question-answer dataset. Four freely available tools—Explore Artificial Intelligence (AI), Cathoven, Questgen, and Lumos Learning—were assessed for this purpose. They were employed to generate a dataset consisting of 40,585 questions with answers derived from the English version of the Quran. The tools exhibited varying levels of performance, with some displaying superior capabilities compared to others. It is noteworthy that although the Cathoven Question Generator had its strengths, the outcomes were not consistent across all tools, highlighting the diverse range of capabilities and results.

In [13], a systematic review of student question generation (SQG) systems is provided. It has been found that since 2000, there has been a growing interest in the development of SQG systems. The review aims to present a comprehensive overview of existing SQG systems. By utilizing a two-dimensional classification scheme, the study identifies the most commonly integrated additional features and design characteristics among 54 SQG learning systems. The research reveals significant variations in design features among SQG systems: the majority are domain-specific, support only one type of question generation, and do not allow the inclusion of multimedia files in student question generation.

In [14], the discipline of AQG, methodologies, datasets, evaluation metrics, and various applications of natural language question generation are discussed. These systems can aid chatbots in generating questions from text, enabling the automatic generation of questions. One potential classification of question generation systems includes visual question generation, stand-alone question generation, and conversation-oriented question generation. The literature presented in the review contributes to understanding the diverse question generation systems and delving into the existing data collections.

The emergence of ChatGPT has sparked intense debate and attention on AI worldwide and has further facilitated the widespread adoption of AI [15], [16], [17], [18], [19], and [20].

In [6], a literature review discusses the impact of ChatGPT in the field of education. ChatGPT is an AI-powered chatbot that was launched in November 2022. It has the capability to generate coherent and informative responses that simulate human conversation. The paper outlines ChatGPT's functionalities across various subjects, its application in education, and the initial challenges faced within the first three months. Through the analysis of 50 articles, it was found that ChatGPT's performance varied, with issues arising from generating inaccurate information and bypassing plagiarism detection systems. The paper recommends that educational institutions promptly revise their evaluation methods and policies and emphasize the importance of providing adequate training to teachers and students for responsible use of ChatGPT in educational environments.

Despite a large body of research and many useful results, there are still areas for further research and unresolved issues in the area of task generation that remain to be addressed. In addition to the indications in the above articles, for example, in [9], the importance of critical thinking and fact-checking is highlighted, and recommendations are made on how to use these models in a responsible and ethical manner in education. The challenges facing the field, such as occasional meaninglessness and the lack of naturalness in the context of information extraction from generated questions, are emphasized in [14]. The biggest problem is that the quality of the questions cannot currently be trusted enough to be used without constant human supervision and regular human review [9].

As can be seen from the above studies, researchers have recently been searching for new and applicable results and answers regarding the utilization of AI [5], [6], [8], [10], [12], and [14]. AI can effectively process lengthy materials and textbooks and generate questions from them. Unfortunately, the quality of AI-generated questions cannot currently be trusted enough to be used without human review. Several studies have also compared different question generation applications, highlighting the strengths and weaknesses of each option. In conclusion, despite significant recent progress in automated question generation, none of the solutions can be considered fully reliable yet. Therefore, it is worthwhile to continue research in this area and leverage the potential of AI tools.

Based on the literature review, it can be assumed that the capabilities of the previously developed Exercise Generation Algorithm+ (EGAL+) could be greatly enhanced by integrating an AI tool for generating questions, given that its scalability has also increased. The goal of EGAL+ was to automate the generation of exams used in educational institutions for assessment, ensuring that they are sufficiently varied, contain tasks that can be included together based on a user-specified set of criteria, and that the exams have the same aggregate difficulty values, ensuring fairness. In a previous literature review, it was found that an application precisely addressing the stated issue does not yet exist, so the significant contribution of EGAL+ is well-founded [21].

Although EGAL+ has proven to be a useful tool, it has several limitations, including (i) restricted input parameters and user parameterization options, (ii) the need for manual question creation by the instructor, and (iii) difficulty in populating the voluminous preference matrix.

This led to the formulation of the following three research objectives for this paper:

i) Enhance the program's logic to improve scalability, enabling it to generate high-quality task sequences from a question bank of at least a thousand questions. This enhancement should maintain similar execution times as with shorter inputs and eliminate previous parameter restrictions.
ii) Integrate an AI tool capable of generating at least a thousand textbook questions, formatted and ready for assembly in EGAL+, thereby eliminating the need for manual question creation.
iii) Enhance support for populating the preference matrix.

This paper presents a novel solution that synergizes modern AI capabilities with a metaheuristic tool. It leverages automated question generation and minimizes the need for human review by processing a whole body of knowledge. This approach makes the questions easily reusable for further high-quality exams with the inclusion of Exercise Generation Algorithm+.

2 THE PURPOSE AND PROCEDURES OF EGAL+

The purpose of EGAL+ is to automate the generation of exams used during assessments in educational institutions. It ensures that the exams are diverse, contain tasks that can be grouped together based on user-specified criteria, and that the task sequences have equal aggregated difficulty values to create fair exercises [21].

The EGAL+ algorithm utilizes a harmony search (HS) metaheuristic approach because of the effectiveness of metaheuristics in solving intricate optimization problems where conventional methods may be suboptimal. In particular, the HS algorithm has shown its capability to efficiently explore the solution space and pinpoint near-optimal solutions for complex problems [22], [23], and [24].

Harmony Search is a metaheuristic optimization algorithm that was initially proposed in 2001. The algorithm is inspired by the way musicians improvise to find pleasing harmony. The idea behind HS is to mimic the behavior that naturally occurs when musicians play their instruments or create music together [25].

Harmony Search, a relatively recent addition to optimization algorithms, has gained widespread usage in various fields for global optimization tasks, owing to its operational simplicity and impressive performance. The HS method is characterized by its efficiency in finding solutions. It has been effectively applied in various fields,

such as mechanical structure design, pipe network optimization, data classification system optimization, and function optimization [26].

The effectiveness of such procedures largely depends on their specific implementation in terms of code. Since they are optimization procedures, numerous algorithms can be developed to yield satisfactory results, with the key distinction among them lying in the requirements for computational power, storage space, and processing time.

Recently, EGAL+ has undergone major improvements through the implementation of mutual dependence conditions on the inputs. This allows anyone without expertise in the field of mathematical optimization to parameterize the program with inputs that lead to a successful run. A run is deemed successful if it produces an output within an acceptable time using the limited resources defined in the given case. In this scenario, all task sequences are equally difficult, and there are no pairs of tasks in any sequence that are prohibited from being included together based on user input. A task sequence is considered of higher quality the more it differs from others and the more it includes tasks that the user has identified as more desirable for inclusion together.

To implement this, the program requires the tasks themselves (*task contents*) as input, from which the task sequences will be assembled. It also needs the difficulty values of the individual tasks on a scale of 1–5 (*task difficulties*) and the preference for the joint inclusion of each pair of tasks on a scale of 0–10 (*coexistence preferences*). Here, 0 represents a prohibitive value, while a higher number indicates a stronger desire for joint inclusion. Additionally, the program needs to know the number of tasks in a task sequence (*exercise length*) and the quantity of task sequences to be generated (*population size*). The combination of task contents and task difficulties will be referred to as the *question bank*.

As the next step, the target difficulty of the task sequences is determined, which is calculated as the sum of the task difficulties included in the individual task sequences. The user is presented with three options: low, medium, and high total difficulty values. Identifying these three options is not a simple task, as it is essential to ensure that, with the selected sum of task difficulties, a population size of task sequences with the exercise length can be created from the provided question bank, taking into account any potential user restrictions on task pairs.

After the user selects the desired target difficulty level for the task sequences, the initial *population* is generated. The initial population comprises the specified number of task sequences, with each sequence containing the designated number of tasks. The total difficulty of each task sequence matches the target difficulty level, which is calculated based on the sum of the difficulties of all tasks within the sequence. Additionally, no task sequence includes any prohibited task pairs. At this stage, it is ensured that a variety of tasks of equal difficulty, meeting all criteria, have been extracted from the question bank. Subsequently, the program proceeds to the quality enhancement module.

As mentioned above, the quality of the task sequences is determined by the differences compared to each other and the coexistence preferences expressed by the user. The higher the coexistence value, the better the quality. To improve the quality of the population, the program utilizes an HS metaheuristic algorithm. Within this framework, the algorithm aims to enhance the quality of the task sequences of the population for a specified number of generations (the default value is 50). During this quality improvement process, the program first generates a random number between 0 and 1. If this number is greater than *HMCR* (default value 0.5), a completely new task sequence is created. If the number is less than or equal to HMCR, one of the existing task sequences is selected and copied, then modified with a *PAR*

probability (default value of 0.2) for a task. It ensures that the task sequence remains valid, whether it is a completely new one or a modified copy of an existing one.

If the program decides to modify one of the existing task sequences, it may not be certain that an element can be found in the question bank within an acceptable short time to be inserted in place of another element in the task sequence or that the replacement is possible at all. In such cases, after a parameterizable time, the program discards the idea of modifying the task sequence and leaves it unchanged.

At this point, the quality improvement attempt can reach one of three states randomly: it can either create a completely new task sequence that fits into the population, copy one without making any changes, or copy one and randomly change one of the tasks included to another appropriate element from the question bank.

Whichever one of the three possibilities arises, the next step of the program is to compare the newly created task sequence with the worst quality of the existing task sequences. If the new task sequence is of better quality than the current worst one, it replaces it with the new one and discards the previous one. If it is not of better quality than the worst one, then it discards the new one.

Regardless of whether a task sequence has been replaced or not, the program compares whether the average quality of the population has increased by at least epsilon (default value: 0.0000000001). If the average quality of the population has not increased by at least epsilon for the failed *epsilon check limit* times (default value: 10) consecutively, the program aborts the execution of new quality improvement iterations and outputs the finished task sequences. Additionally, if the generation limit is reached, the program concludes the quality improvement iterations and outputs the finished task sequences.

3 IMPROVEMENTS REGARDING EFFICIENCY

A major shortcoming of EGAL+ has been its inability to handle the necessity of creating long task sequences for many students from a question bank with a large number of elements. This is a critical situation where teachers could benefit the most from automation due to the complexity of the task. Prior to the advancements introduced in this study, the program was severely limited, with a maximum population size of 100, a question bank of 50, and a highly restricted number of zero values in coexistence preferences. These limitations constrained the input elements and user parameterization significantly.

One of the objectives of these most recent developments was to eliminate all input size restrictions. This allows users to enter an extensive question bank, potentially in the thousands, created from a large preexisting knowledge base, and generate high-quality task sequences, as previously defined, for a substantial number of students, even numbering in the hundreds.

The ability to question the fundamentals of a field of science is particularly beneficial in an educational setting. This is especially true when these fundamentals either remain constant or change only very slowly, requiring in-depth examination by many students on a regular basis. This scenario is commonly observed in undergraduate university education across various disciplines like mathematics, micro- and macroeconomics, anatomy, law, and history. Moreover, with the rise of online education, the capacity to engage multiple students in questioning simultaneously is further amplified.

To eliminate the input restrictions, one of the most important modifications implemented in the program was to change the operating logic during the initial

population generation. As described in Section 2, EGAL+ first creates an initial population based on the specified parameters, which is valid in terms of the requirements set by the coexistence preferences and the aggregated difficulty goal. So far, the generation of the initial population has been characterized by the fact that the algorithm selected the exercise length and number of elements in a single operation from all the possible tasks in the question bank for a task sequence and then checked whether it was valid to include these tasks together. It is easy to see that the more prohibitive values are in the coexistence preferences, and the more different the individual difficulty values given for each task are, the less likely it is that the correct result will follow from selecting the tasks to be inserted into the task sequence at the same time, because these factors all reduce the number of acceptable combinations.

For this reason, the program was modified so that, during the creation of the initial population, the exercise length and number of tasks are not selected all at once. Instead, the selection process is organized into iterations to choose only one new task at a time that meets the requirements set by the coexistence preferences. This process continues until enough tasks are successfully chosen for a task sequence.

Even in this solution, additional control operations are required. For example, preparations must be made for the possibility that occasionally a task is chosen that, although it corresponds to the ones chosen up to that point, does not allow the completion of the whole task sequence. In this case, it must be ensured that the program does not enter an infinite loop. Overall, a better result can be achieved with this new approach than with the previously used method.

Although at first it may seem like a more computationally demanding task to check the task sequence each time after a component is added to see if it is built correctly, it still results in a much more efficient run than checking it once after it has been created. This is because it greatly reduces the probability of creating unnecessary task sequences, thereby reducing the overall time required for generating the output.

The method of storing population elements has also been changed. Previously, individual elements were represented as 0s and 1s in a vector, where a value of 1 indicated an included task and a value of 0 indicated a non-included task. This resulted in a vector with a length equal to the number of questions in the question bank. Now, instead of using 0s and 1s, the indexes of the included tasks are stored. This change reduces the size of population elements and streamlines processing, as the length of a task sequence is now determined by the exercise length rather than the potentially larger size of the question bank.

The third significant modification concerning the efficiency of the program is that it now generates the starting population while searching for target difficulty options. Instead of initially finding the potential population target difficulty options in a separate operation and then creating the initial population after the user's selection, the difficulty options are now determined by analyzing randomly generated task sequences and classifying them. This process results in creating an initial population for each possible target difficulty option. After the selection, this initial population can be treated as such, allowing the program to commence enhancing its quality. Consequently, this approach significantly reduces the overall operational requirements.

With the help of the developments outlined in this chapter, the need for input size limitations has been successfully eliminated. It is now possible to generate exams for hundreds of students from a question bank containing thousands of questions in just seconds of execution time, without any solvable coexistence prohibitions. In the previous state of the program, such an attempt would have led to practically infinite running.

For the exact implementation, the source code of EGAL+ can be found in the following GitHub repository: https://github.com/balazs-domsodi/EGALPP.

4 INTEGRATING GENERATIVE AI FOR STREAMLINING QUESTION BANK CREATION

Although recent developments have significantly enhanced the program's scalability, as discussed in Section 3, these advancements have also introduced new challenges that must be addressed before EGAL+ can be deployed in a real-world setting. One of these challenges is that the creation and categorization of numerous questions to fully leverage the program's benefits may exceed instructors' resources in most scenarios.

Given the substantial challenge in this case of creating a large volume of questions in textual form based on an extensive text source, it was identified that this issue can be addressed through the integration of an AI module to generate questions from a knowledge base, such as a textbook. Subsequently, the recently described metaheuristic algorithm can compile the specific exercises, providing a comprehensive solution for generating exams, which is the primary purpose of Exercise Generation Algorithm+.

In the field of natural language processing, AQG applications are prominent. These applications autonomously generate questions from text sources and images, guided by a particular subject or concept. They have been extensively tested in educational environments and are regularly used in machine reading comprehension tools and conversational systems. This has led to a significant increase in their popularity due to their versatility and effectiveness. Closed- and open-domain AQG applications can be distinguished. For closed-domain question generation, queries are formulated specific to a field such as medicine or educational literature, drawing upon knowledge that is unique to the domain and bound by an ontology. In contrast, open-domain question generation is not tied to any specific domain and permits the creation of questions regardless of the domain, necessitating only global ontologies [14]. Because the goal of this research is to create an application for general use, open-domain options could be considered.

The utilization of LLMs for AQG has undergone significant development in recent years, evolving from early approaches that treated MCQ generation as a pipeline of subtasks to more recent deep learning approaches. These models are now capable of generating questions, answers, and distractor answer options for MCQs by providing them with sentences from a textbook. Despite some challenges, studies have shown that LLMs are remarkably capable of creating MCQs matching human performance on most metrics [10]. Considering these recent findings and the necessity for high-quality questions in the EGAL+ question bank that can replace human-generated content, along with the potential for future automatic evaluation, the objective was to incorporate an AQG solution using LLM for generating multiple-choice questions.

For selecting the appropriate tool in a specific context of generating MCQs, research studies are actively comparing various alternatives to LLMs such as GPT-3 and GPT-4, as well as different AQG applications. These studies measure their effectiveness in generating relevant and accurate questions and answers for MCQs. The findings suggest that while some tools outperform others, there is considerable potential in this field [5] [12].

Given the availability of numerous suitable tools, it could be beneficial to incorporate multiple alternatives into EGAL+ in the future. This would allow users to select the application programming interface that best fits their specific use case. The primary objective of this research is to demonstrate that an AI-based question generation tool can be effectively combined with a metaheuristic algorithm for mass-producing high-quality exams. To illustrate this concept, PrepAI is selected and integrated, a popular tool available at https://prepai.io. Notably, PrepAI meets all the previously established criteria for a tool suitable for implementation in EGAL+, known for its versatility and ability to process PDF inputs.

5 PRACTICAL DEMONSTRATION OF RESULTS

In this section, the capabilities of the recently improved EGAL+ are demonstrated, particularly its efficiency in integrating an AI module for processing numerous questions and describing each of the program's three main modes. The current UI of the program listing its modes is shown in Figure 1.

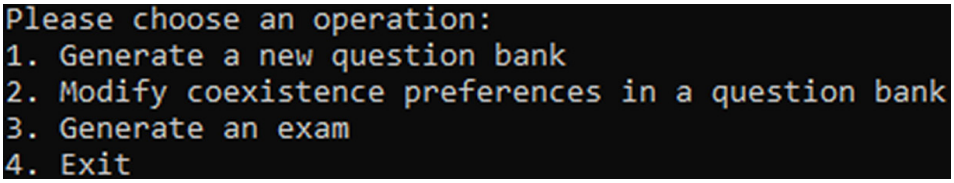

Fig. 1. Screenshot showing the current UI of the program, listing its modes in a menu structure

The first step of the program is to generate new question banks. This process requires providing a name for the question bank to be created, a topic name, the input file (typically a textbook), the number of questions to be generated, and the page count and/or specific page ranges to be included from the input file for question generation.

Thanks to the advanced capabilities of PrepAI, it is now possible to import hundreds of pages from a PDF document and generate up to a thousand questions. Additionally, due to its seamless integration with EGAL+, it can be operated directly from within this program. This integration guarantees that the resulting text file containing the question bank is already formatted for further processing in EGAL+. Each row in this file includes the generated question, the suggested difficulty value for the question (ranging from 1–2 as per PrepAI), and coexistence preferences ranging from 0 to 10, separated by semicolons. Each coexistence preference value indicates the desired joint inclusion of that question with the preceding ones in order, with the default value being 10 at the time of question generation. The answer options are also separated by semicolons, with the correct answer denoted by an asterisk. All four elements in each row are separated by tab characters. The parameterization of the first mode can be seen in Figure 2.

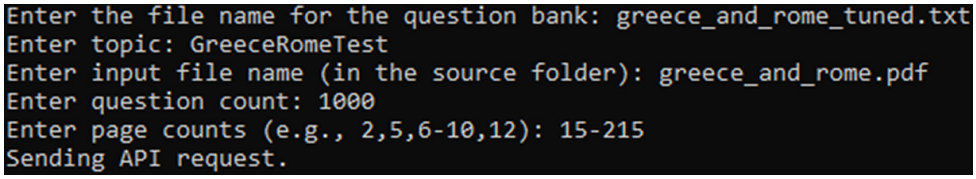

Fig. 2. Screenshot showing an example parameterization of the first mode

The second mode is a new improvement aimed at enhancing the support for populating the preference matrix by allowing the optional modification of coexistence preferences. This mode requires the question bank itself as an input. Users can define as many question groups as they desire by specifying the question numbers and/or ranges, along with the intended coexistence preferences for each group. Once the user finishes defining the question groups, they are prompted to provide coexistence preferences for the disjoint ones from the created groups. Users are asked to specify the coexistence preference values for each pair, and the corresponding values in the question bank are then adjusted based on the user input. The parameterization of the second mode can be seen in Figure 3.

```
Enter the file name of the question bank: greece_and_rome_tuned.txt
Enter the group of questions (e.g., 2,5,6-10,12): 1-38
Enter the joint inclusion preference (0-10): 6
Do you want to define another group of questions? (yes/no): yes
Enter the group of questions (e.g., 2,5,6-10,12): 39-140
```

Fig. 3. Screenshot showing an example parameterization of the second mode

The generation of the exams is the third and final mode, which involves HS metaheuristic optimization. The program takes the generated question bank with optionally modified coexistence preferences as input. It also prompts the user for the number of desired questions in an exam and the number of exams to be generated. The program then identifies three achievable total difficulty goals for the exams with the largest possible difference between them to provide the user with diverse options. Once the desired difficulty goal is selected, the program generates the exams. It considers all coexistence preferences, potential prohibitions on including specific questions, strictly maintains the difficulty goal, and aims to create the most varied exams possible while adhering to all criteria. The parameterization of the third mode can be seen in Figure 4.

```
Enter the name of the question bank: greece_and_rome_tuned.txt
Please specify the desired exercise length:
30
Please specify the desired population size:
100
Please choose from the difficulty options below:
67
83
99
```

Fig. 4. Screenshot showing an example parameterization of the third mode

For this demonstration, a question bank consisting of 1000 questions was generated from the first 200 pages of the book “Myths and Legends of Ancient Greece and Rome” by E. M. Berens [27] using the program’s first mode. This book was accessed through a volunteer-led initiative dedicated to the preservation and digitization of culturally significant works known as Project Gutenberg, accessible at https://gutenberg.org. The topic selection was guided by its inclusion in general education to ensure the broad interpretability of the demonstration.

For the generated 1000 questions, the following question groups were created in the second mode: G1: 1–38, G2: 39–140, G3: 141–279, G4: 280–540, G5: 541–635, G6: 636–1000, and G7: 637–649. The coexistence preferences in each group were set to 6, except for G6, where it was set to 3, and G7, where it was set to 0. The coexistence preferences between disjoint groups were set to 10 in each case, except for G4–G5, where they were set to 0, and for G6, which for each disjoint other group was set to 7.

The coexistence preference values used for the questions derived from the book are randomly selected rather than expertly determined. However, this randomness does not impact the successful operation of the program.

To demonstrate the ability of EGAL+ to handle difficulty values for single questions in the range of 1–5, the initial value of 1 for PrepAI in the question bank was randomly adjusted within the range of 1–2, and the initial two values were randomly adjusted within the range of 3–5.

In the third mode, 100 exams of 30 questions were generated by selecting the medium difficulty option out of the three total difficulty options. The program’s quality improvement module can enhance the results by approximately 1.3% (as detailed in Chapter 4), surpassing the initial population that already meets all strict criteria.

It is notable that after generating 1000 questions from the 200 pages using the integrated PrepAI module, the generation of 100 exams of 30 questions was completed in less than a second by the HS module of the program, meeting all the specified user criteria. This showcases the power and efficiency of EGAL+ when integrated with an AI module to produce high-quality exams rapidly, requiring minimal manual effort from the users.

All the results in this demonstration were obtained from a laptop equipped with an Intel Core i7-9750H 2.6 GHz CPU and 16.0 GB of RAM. The resulting question bank is available at https://github.com/balazs-domsodi/EGALPP/blob/main/databank/greece_and_rome.txt, and the resulting task sequences are available at https://github.com/balazs-domsodi/EGALPP/blob/main/output/tasks.txt. The task indices of the initial population are available at https://github.com/balazs-domsodi/EGALPP/blob/main/output/initial.txt, and the task indices of the enhanced population are available at https://github.com/balazs-domsodi/EGALPP/blob/main/output/enhanced.txt, with the chosen aggregated difficulty value in the beginning of both files.

It's important to note that while the modes for adjusting coexistence preferences and generating exams only require suitable inputs, creating new question banks necessitates a PrepAI account. Users must possess such an account and input the required credentials into a client_id.txt and client_secret.txt file located in the program's root directory.

This study presents a method for educators to automatically process textbooks and generate question banks, enabling the creation of high-quality exams in seconds. However, it's important to note that the AI used for question generation may introduce errors. The authors propose an initial, one-time investment of time for an expert to review and correct every AI-generated question bank. This step is crucial until generative AI achieves full professional reliability. The primary advantage of EGAL+ is that once the generated question bank is proofread and the coexistence preference values are fine-tuned, educators can repeatedly generate high-quality exams from specific course materials for an extended period of time in subjects that don't change rapidly. This process ensures long-term, consistent quality in assessments, freeing up educators' time spent assembling exams from textbooks.

6 CONCLUSION AND FUTURE IMPROVEMENTS

This study addresses a practical problem in educational institutions: the manual compilation of exercises often leaves educators with less time for tasks that require their personal attention, such as providing individualized student support. Alternatively, if insufficient time is allocated to exercise preparation, the quality of these materials may be compromised, potentially hindering optimal student assessment.

The research problem identified is the computational automation of exercise assembly from educational resources. This complex problem presents numerous challenges, including ensuring the quality and appropriateness of the assembled exercises and handling diverse educational resources.

A review of the literature revealed that while many sophisticated exercise-generating systems exist, none are capable of addressing the specific problem identified in this study: generating multiple, distinct subsets of predetermined tasks based on quality criteria such as difficulty and categorization. This represents a research gap, as none of the solutions or ongoing research discovered by the authors aim to address exercise generation and compilation in such a generic and standardizable manner.

The objectives of this research were formulated as follows:

i) Enhance the program's logic to improve scalability, enabling it to generate high-quality task sequences from a question bank of at least 1000 questions. This enhancement should maintain similar execution times as with shorter inputs and eliminate previous parameter restrictions.
ii) Integrate an AI tool capable of generating at least 1000 textbook questions, formatted and ready for assembly in EGAL+, thereby eliminating the need for manual question creation.
iii) Enhance support for populating the preference matrix.

The research objectives were successfully met by enhancing the program's logic and integrating a generative AI tool into the program's operations, as documented and demonstrated in the paper. To achieve the research objectives, the authors further developed an algorithm they had previously created for generating task sequences, Exercise Generation Algorithm+.

The improvements presented in this study focus on enhancing the program's scalability to process more data and integrating an AI-based text-processing module capable of generating a large number of MCQs from PDF-based sources. The study demonstrates and documents the effectiveness and efficiency of the resulting program, indicating that all the research objectives were successfully achieved. All information necessary to reproduce the results is made available.

The enhanced EGAL+ has proven capable of automatically processing textbooks and generating questions, a feat made possible by the integrated AI module. Following a single proofreading session and fine-tuning the coexistence preference values, educators can leverage the implemented metaheuristic algorithm to consistently generate high-quality exams from the provided course materials in a remarkably short time. This process ensures the long-term, consistent quality of assessments and reduces the time teachers spend compiling exams from textbooks.

However, there are still many opportunities for further development of the program. One of the most significant opportunities is updating difficulty values based on the results achieved by participants on the prepared assessments. This would enable the question bank to have increasingly accurate difficulty values after each use, facilitating the self-improvement of the question bank through usage. Another potential enhancement for the future involves integrating the AI tool's generated answer options into the program and implementing an automatic assessment feature to provide additional support to instructors. Other potential improvements include exploring integration possibilities with learning management systems and designing a user interface that caters to the needs of educators.

7 REFERENCES

[1] V. Nentwich, N. Fischer, A. C. Sonnenbichler, and A. Geyer-Schulz, "Computer aided exercise generation – A framework for human interaction in the automated exercise generation process," in *Proceedings of the 13th International Joint Conference on e-Business and Telecommunications (ICETE 2016)*, 2016, pp. 57–63. https://doi.org/10.5220/0005947700570063

[2] Y.-A. Bachiriand and H. Mouncif, "Artificial intelligence system in aid of pedagogical engineering for knowledge assessment on MOOC platforms: Open EdX and Moodle," *International Journal of Emerging Technologies in Learning (iJET)*, vol. 18, no. 5, pp. 144–160, 2023. https://doi.org/10.3991/ijet.v18i05.36589

[3] S. Pandraju and S. G. Mahalingam, "Answer-aware question generation from tabular and textual data using T5," *International Journal of Emerging Technologies in Learning (iJET)*, vol. 16, no. 18, pp. 256–267, 2021. https://doi.org/10.3991/ijet.v16i18.25121

[4] N. Willert and J. Thiemann, "Template-based generator for single-choice questions," *Technology, Knowledge and Learning*, vol. 29, pp. 355–370, 2024. https://doi.org/10.1007/s10758-023-09659-5

[5] A. Tran, K. Angelikas, E. Rama, C. Okechukwu, D. H. Smith, and S. MacNeil, "Generating multiple choice questions for computing courses using large language models," in *2023 IEEE Frontiers in Education Conference (FIE)*, College Station, TX, USA, 2023, pp. 1–8. https://doi.org/10.1109/FIE58773.2023.10342898

[6] C. K. Lo, "What is the impact of ChatGPT on education? A rapid review of the literature," *Education Sciences*, vol. 13, no. 4, p. 410, 2023. https://doi.org/10.3390/educsci13040410

[7] C.-Y. Chung, I.-H. Hsiao, and Y.-L. Lin, "AI-assisted programming question generation: Constructing semantic networks of programming knowledge by local knowledge graph and abstract syntax tree," *Journal of Research on Technology in Education*, vol. 55, no. 1, pp. 94–110, 2022. https://doi.org/10.1080/15391523.2022.2123872

[8] A. Ushio, F. Alva-Manchego, and J. Camacho-Collados, "A practical toolkit for multilingual question and answer generation," in *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics*, *System Demonstrations*, Toronto, Canada. Association for Computational Linguistics, vol. 3, 2023, pp. 86–94. https://doi.org/10.18653/v1/2023.acl-demo.8

[9] E. Kasneci *et al.*, "ChatGPT for good? On opportunities and challenges of large language models for education," *Learning and individual differences*, vol. 103, p. 102274, 2023. https://doi.org/10.1016/j.lindif.2023.102274

[10] A. M. Olney, "Generating multiple choice questions from a textbook: Llms match human performance on most metrics," in *AIED Workshops*, 2023.

[11] J. Heilala, A. Shibani, and A. Gomes de Freitas, "The requirements for heutagogical attunement within STEAM education," *International Journal of Emerging Technologies in Learning (iJET)*, vol. 18, no. 16, pp. 19–35, 2023. https://doi.org/10.3991/ijet.v18i16.42313

[12] S. S. M. Alnefaie, E. Atwell, and M. A. Alsalka, "Using automatic question generation web services tools to build a Quran question-and-answer dataset," *International Journal on Islamic Applications in Computer Science and Technology*, vol. 11, no. 2, pp. 1–12, 2023.

[13] F.-Y. Yu and C.-W. Kuo, "A systematic review of published student question-generation systems: Supporting functionalities and design features," *Journal of Research on Technology in Education*, vol. 56, no. 2, pp. 172–195, 2022. https://doi.org/10.1080/15391523.2022.2119448

[14] N. Mulla and P. Gharpure, "Automatic question generation: A review of methodologies, datasets, evaluation metrics, and applications," *Progress in Artificial Intelligence*, vol. 12, pp. 1–32, 2023. https://doi.org/10.1007/s13748-023-00295-9

[15] H. H. Thorp, "ChatGPT is fun, but not an author," *Science*, vol. 379, no. 6630, pp. 313–313, 2023. https://doi.org/10.1126/science.adg7879

[16] E. A. Van Dis, J. Bollen, W. Zuidema, R. van Rooij, and C. L. Bockting, "ChatGPT: Five priorities for research," *Nature*, vol. 614, no. 7947, pp. 224–226, 2023. https://doi.org/10.1038/d41586-023-00288-7

[17] A. Tatnall and A. Fluck, "Twenty-five years of the education and the information technologies journal: Past and future," *Education and Information Technologies*, vol. 27, pp. 1359–1378, 2022. https://doi.org/10.1007/s10639-022-10917-9

[18] T. T. A. Ngo, "The perception by university students of the use of ChatGPT in education," *International Journal of Emerging Technologies in Learning (iJET)*, vol. 18, no. 17, pp. 4–19, 2023. https://doi.org/10.3991/ijet.v18i17.39019

[19] O. Abu Khurma, N. Ali, and R. Hashem, "critical reflections on ChatGPT in UAE education: Navigating equity and governance for safe and effective use," *International Journal of Emerging Technologies in Learning (iJET)*, vol. 18, no. 14, pp. 188–199, 2023. https://doi.org/10.3991/ijet.v18i14.40935

[20] L. H. Al-Obaydi, M. Pikhart, and B. Klimova, "ChatGPT and the general concepts of education: Can artificial intelligence-driven chatbots support the process of language learning?" *International Journal of Emerging Technologies in Learning (iJET)*, vol. 18, no. 21, pp. 39–50, 2023. https://doi.org/10.3991/ijet.v18i21.42593

[21] B. Láng and B. Dömsödi, "Development of the improved exercise generation metaheuristic algorithm EGAL+ for end users," *International Journal of Emerging Technologies in Learning*, vol. 17, no. 11, pp. 210–224, 2022. https://doi.org/10.3991/ijet.v17i11.28099

[22] M. Dubey, V. Kumar, M. Kaur, and T. P. Dao, "A systematic review on harmony search algorithm: Theory, literature, and applications," *Mathematical Problems in Engineering*, vol. 2021, no. 1, p. 5594267, 2021. https://doi.org/10.1155/2021/5594267

[23] F. Qin, A. M. Zain, and K. Q. Zhou, "Harmony search algorithm and related variants: A systematic review," *Swarm and Evolutionary Computation*, vol. 74, p. 101126, 2022. https://doi.org/10.1016/j.swevo.2022.101126

[24] M. Mahdavi, M. Fesanghary, and E. Damangir, "An improved harmony search algorithm for solving optimization problems," *Applied mathematics and computation*, vol. 188, no. 2, pp. 1567–1579, 2007. https://doi.org/10.1016/j.amc.2006.11.033

[25] Z. W. Geem, J. H. Kim, and G. V. Loganathan, "A new heuristic optimization algorithm: Harmony search," *Simulation*, vol. 76, no. 2, pp. 60–68, 2001. https://doi.org/10.1177/003754970107600201

[26] X. Z. Gao, V. Govindasamy, H. Xu, X. Wang, and K. Zenger, "Harmony search method: Theory and applications," *Computational intelligence and neuroscience*, vol. 2015, no. 1, p. 258491, 2015. https://doi.org/10.1155/2015/258491

[27] E. M. Berens, "Myths and Legends of Ancient Greece and Rome," Urbana, Illinois: Project Gutenberg. Retrieved from https://www.gutenberg.org/ebooks/22381

8 AUTHORS

Blanka Láng is an Associate Professor at the Corvinus University of Budapest. Her research areas include harmony and genetic metaheuristic algorithm development, linear regression model selection, and exercise generation problems (E-mail: blanka.lang@uni-corvinus.hu).

Balázs Dömsödi is a Ph.D. student in Business Informatics Engineering at Corvinus University of Budapest. His research interests include harmony metaheuristic algorithm development and exercise generation problems (E-mail: balazs.domsodi@stud.uni-corvinus.hu).

Online-Journals.org

iJET International Journal of Emerging Technologies in Learning

iJET | eISSN: 1863-0383 | Vol. 19 No. 6 (2024) | OPEN ACCESS

https://doi.org/10.3991/ijet.v19i06.48343

PAPER

A Virtual Reality Game to Promote the Role of a Healthy Diet in Male Reproduction

Elena Bolognesi[1], Elena Moretti[2], Maurizio Masini[3], Cinzia Signorini[2], Vincenzo Santalucia[3], Giulia Collodel[2](✉)

[1]Freelance dietician, Italy

[2]Department of Molecular and Developmental Medicine, University of Siena, Siena, Italy

[3]Department of Social, Political and Cognitive Sciences, University of Siena, Siena, Italy

giulia.collodel@unisi.it

ABSTRACT

This paper presents gaming as an alternative way of transmitting information compared to traditional academic methods. New technologies provide an opportunity, and in particular, virtual reality gaming allows users to have an immersive, personifying, intuitive, and engaging experience useful in visualizing complex concepts. Questionnaires were administered to secondary school students to extrapolate relevant information regarding the role of diet on sperm quality following the experience of the virtual reality game "Oxistress." This game allows participants to understand in a playful and recreational way how the presence of reactive oxygen species can damage spermatozoa and how adequate diet choices provide ammunition in the form of antioxidants and omega-3 fatty acids to protect and repair spermatozoa. The results suggest that virtual reality (VR) may represent an important way to improve knowledge of scientific topics. Further studies will be necessary, involving larger populations with different basic knowledge as well as checks on whether the information acquired remains in memory after months.

KEYWORDS

diet, learning, omega 3, reactive oxygen species, sperm quality, virtual reality (VR)

1 INTRODUCTION

Virtual reality (VR) is an innovative tool that, due to its immersive and multisensory nature, can fulfill the principles of active learning. The immersive experiences it offers and the sense of presence and embodiment are all key factors that can promote knowledge acquisition [1]. VR offers three main opportunities: it can convert the abstract into the tangible, it supports doing rather than just observing, and it can replace learning methods that would be ideal but have little application in reality [2]. Especially when applied to studies related to health and nutrition education [3], we can say that games can be equally, if not more, effective than traditional learning methods [4].

Bolognesi, E., Moretti, E., Masini, M., Signorini, C., Santalucia, V., Collodel, G. (2024). A Virtual Reality Game to Promote the Role of a Healthy Diet in Male Reproduction. *International Journal of Emerging Technologies in Learning (iJET)*, 19(6), pp. 52–64. https://doi.org/10.3991/ijet.v19i06.48343

Article submitted 2024-02-02. Revision uploaded 2024-04-03. Final acceptance 2024-04-03.

In recent years, various studies have shown a decrease in sperm quality, which can be linked to increased exposure to chemicals such as pesticides, plastics, and other anthropogenic materials [5]. Additionally, lifestyle factors such as smoking habits [6] and/or excessive alcohol consumption can affect sperm production and quality [7]. Over the last two decades, sperm concentration has exhibited a further progressive decline, and a significant decrease in fertilization rates has been hypothesized over the next 30 years [8].

Dietary patterns, the components of diet, and nutrients have been studied as possible determinants of sperm function and fertility [9, 10]. Anti-inflammatory diets, such as the Mediterranean diet (which includes increased intake of monounsaturated and n-3 polyunsaturated fatty acids, flavonoids, and reduced intake of red and processed meat), have been shown to improve fertility, assisted reproductive technology (ART) success, and sperm quality in men [11].

In men, the ability to modulate inflammatory processes and reactive oxygen species (ROS) formation through dietary and non-pharmacological treatments may be essential to improving male reproductive outcomes. Unfortunately, humans cannot serve as an ideal model for study due to the significant variability in genetics, behavior, and lifestyles. Animal models can be beneficial in research.

Our group obtained scientific data on the effect of an omega-3 diet on rabbit sperm and testes [12]. This data was used to conduct a VR experiment.

The VR game "Oxistress," developed by the Virtual Reality Laboratory (LabVR) of the Department of Political, Social, and Cognitive Sciences in collaboration with the Department of Molecular and Developmental Medicine at the University of Siena, has made it possible to visualize the impact of proper nutrition on spermatozoa. The game demonstrates, in a playful and recreational manner, how the presence of ROS can harm sperm and how making appropriate dietary choices can supply antioxidants and omega-3 fatty acids to safeguard and restore sperm. The concept is to educate through interactive gameplay [13].

The aim of this study was to assess the ability of a group of students to acquire scientific information while using Oxistress.

2 MATERIALS AND METHODS

2.1 Participants

The application of virtual reality was tested on a sample of 30 secondary school students (aged 17–19) from the Technical Technological Institute and High School of Applied Sciences "Tito Sarrocchi" in Siena. The sample consisted of 14 males and 16 females. One female student was excluded from the study due to anxiety issues related to the experience. Prior to participation, the students' parents provided consent for their children to take part in the project. All anti-Covid procedures recommended by the Prevention and Protection Service of the University of Siena were adhered to.

2.2 Experimental protocol

The experience took place in March 2023 at the Santa Chiara Lab of the University of Siena in LabVR. Two areas were arranged for the experiment. The first area was furnished with four VR game stations, each equipped with the Oculus Quest 2 visor (Meta, Menlo Park, California, United States), and a computer or tablet for researchers to assist students during the simulation by observing the game. The second area was designated for completing questionnaires regarding the experience.

Before the VR session began, a brief introductory talk was given about the global trend of male infertility in the world and the VR experience that the students would undergo. Subsequently, each participant was led through the game simulation. It took between 30 and 40 minutes to complete the entire simulation.

During the game, it is possible to open information panels on the various biological topics covered.

Oxistress is a narrative game designed to educate a non-specialized audience about the biological process of oxidative stress and the roles of free radicals, antioxidants, and fatty acids in either accelerating or limiting it. By engaging with the game, players can understand the impact of dietary habits on the system, especially on the quality of the seminal fluid, and overall reproductive health.

In Oxistress, the player helps a group of astronauts make optimal dietary choices to maintain their fertility levels and ensure the success of the colonization mission.

The science-fiction setting provides a narrative strategy to isolate the system you want to explain and maximize agency, creating a sensible context for the association of individual game choices (such as what to eat) with immediate macroscopic consequences (such as the effect on the fertility of the crew).

The game was developed by the LabVR UNSI team using the Unity game development engine and runs on modern VR headsets such as Meta Quest. Here follows a link to a short video introducing the VR game: https://www.youtube.com/watch?v=JgRwJ4UaP7E

Two questionnaires were designed to evaluate the students' learning. They were completed anonymously, one before playing the game and one after the simulation was conducted. The pre-experience questionnaire aimed to understand the participants' initial knowledge in the fields of nutrition and reproductive biology (see Figure 1). In contrast, the post-experience questionnaire aimed to assess the degree of specific knowledge achieved and new information acquired in the fields of reproductive biology and nutrition through specific questions about what had been presented in the game (see Figure 2). Additionally, LabVR researchers administered a questionnaire to evaluate the virtual reality simulation itself. This aimed to understand the level of engagement, enjoyment of the game, and any feelings of dizziness or nausea to enhance the Oxistress game.

1. On a scale of 0 to 10 how much attention do you pay to your food choices?

0	1	2	3	4	5	6	7	8	9	10

2. On a scale of 0 to 10 how much do you think food affects your overall health?

0	1	2	3	4	5	6	7	8	9	10

3. Of these, which do you think are the principles of the Mediterranean diet?

- Consumption of fruit, vegetables, legumes, whole grains, dried fruit, extra virgin olive oil and fish
- Consumption of fruit, vegetables, poultry and red meat, dairy products, restricted bread and pasta
- Consumption of refined cereals, butter, margarine, few fruit and vegetables, preference of eggs and cold cuts as main course

4. Which of these foods do you think should be preferred because of its nutritional properties?

- Fish
- Red meat
- Matured cheeses (parmesan, pecorino, grana padano...)

5. Have you ever heard of saturated and trans fats?

- Yes
- No

6. Do you believe that your eating habits can affect your reproductive capacity?

- Yes
- No

7. Have you ever studied the spermatozoon?

- Yes
- No

8. Which of these statements do you think is correct?

- The spermatozoon consists of a part containing the nucleus called the head, an intermediate portion and a tail called the flagellum which allows movement
- They have a diploid genetic make-up, meaning they have two copies of genetic material for each chromosome
- They have a head containing the genetic material surrounded by a protein membrane

9. What do you think is a free oxygen radical (ROS)?

- Molecules with a stray electron produced by the body due to processes such as oxidative stress or normal oxidation-reduction reactions in the cell
- Highly reactive molecules involved in repairing cell damage
- Totally harmless waste products of the cell

10. Have you ever heard of antioxidants?

- Yes
- No

11. Given the proposed statements, what do you think is the correct definition of antioxidants?

- They are substances found inside the cell and that we can take in through food, which are able to inhibit oxidation processes that damage cellular material
- They are enzymes that enable the digestion of food containing oxygen molecules
- They are substances exclusively produced by the cell to fight the damage caused by the waste products of cell metabolism

12. Have you ever heard of omega-3?

- Yes
- No

13. In your opinion, which group of macromolecules do omega-3s belong to?

- Fats
- Protein
- Carbohydrates

Fig. 1. Pre-experience questionnaire

1. **Based on your experience, which of the following do you think is the correct definition of a free radical?**
   - These are highly reactive molecules involved in repairing damage to the phospholipid membrane of the spermatozoa
   - Waste products of the cell that help counteract the production of antioxidants
   - They are molecules with an unbalanced electron produced by our organism due to processes such as oxidative stress or normal oxidation-reduction reactions of the cell that damage the phospholipid membranes of spermatozoa

2. **Which of these are the parameters that make it possible to understand if the spermatozoon is healthy and efficient?**
   - Vitality, morphology, motility
   - Strength, antioxidant content, head size
   - Tail length, vitality and potency

3. **Based on your experience, to counteract free radicals, what would you try to assume through your diet?**
   - Protein
   - Antioxidants
   - Fatty acids

4. **Based on your recent experience, which of the following can be defined as antioxidants?**
   - Nitrosamines, heterocyclic amines and ethanol
   - Insulin, glucagon and somatostatin
   - Carotenoids, polyphenols, flavonoids, tocopherols and vitamin C

5. **Why are fatty acids relevant to sperm health?**
   - Because they represent the majority of the spermatozoon's plasma membrane
   - Because they increase tail motility
   - Because they counteract the action of free radicals

6. **Based on your experience, which of the following are the most important fatty acids for sperm health?**
   - Monounsaturated fatty acids
   - Polyunsaturated fatty acids (omega 3 and omega 6)
   - Saturated fatty acids

7. **Where are omega 3 fatty acids found most?**
   - Poultry and olive oil
   - Eggs, milk and milk products (yoghurt, fresh cheese, mature cheese)
   - Oily fish, dried fruit and green leafy vegetables

Fig. 2. Post-experience questionnaire

3 RESULTS

The aim of the questions in the pre-experience questionnaire was to assess preliminary and general knowledge about nutrition and reproductive biology. The first two questions required students to rate, on a scale of 0 to 10, the level of attention they personally devote to their food choices (question 1) and their belief in how much their health depends on a balanced diet (question 2).

The score that was most selected for the first question was 7, chosen by eleven students, while only one student gave a score of 10. Only two students were below the sufficient mark, indicating their attention to their food choices as 4. The total arithmetic mean of the scores is 7.32.

In the second question, however, most of the students (n = 25) rated the importance of nutrition for good general health with a score of at least 8. No one rated it lower than sufficient, and only one student marked 6. The total arithmetic mean of the scores was 8.61.

The third question aimed to find out what the students knew about the Mediterranean diet. 62% of the students answered correctly (consumption of

fruit, vegetables, legumes, whole grains, dried fruit, extra virgin olive oil, and fish). A good percentage of students, 28%, thought it was correct to choose the option stating that one of the principles of the Mediterranean diet is to reduce bread and pasta.

In the fourth question, 69% of the students selected fish as the food with the best nutritional properties, compared to 20% who selected red meat and 11% cheese. Question four, as well as questions five and six, were not asked to obtain an unequivocal truth about food, as this does not exist. Instead, one must speak of a varied diet rather than "food" with the best nutritional properties. These questions were designed to assess the students' level of previous knowledge on the subject. For example, the risk factor that red meat, when processed or consumed in excess, may represent for the development of cancer, or whether they were aware that mature cheeses, while rich in proteins of high biological value and minerals such as calcium, are also high in salt and saturated fatty acids and should therefore be limited in their daily intake. In the fifth question, students were asked whether they had ever heard about saturated fatty acids and trans-fatty acids; 83% said yes and 17% said no. To conclude the section on preliminary knowledge of nutrition, the sixth question asked whether there was, in their opinion, a correlation between eating habits and reproductive capacity.

The answer was yes in 79% of the cases and no in 21% (see Figure 3). The questions were then more specific and related to the topics covered in the game. These questions aimed to better understand the role of the virtual reality experience in increasing their knowledge. In question 7, which asked if students had ever studied the spermatozoon, 65% said yes, while 35% said no.

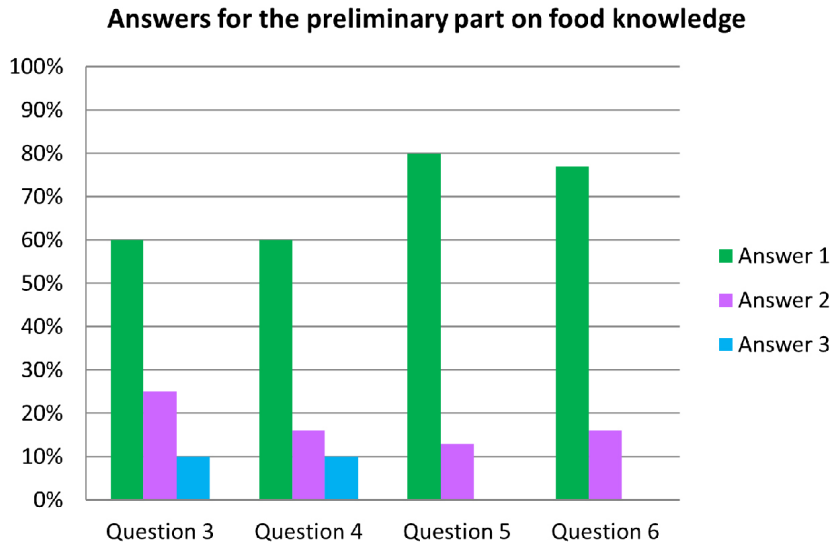

Fig. 3. Histogram showing the percentage of answers given to the questions on food knowledge in the pre-experience questionnaire

Students were then asked to choose one correct affirmation from the statements presented regarding the spermatozoon. 89% of the students selected the correct answer, while 11% gave one of the two incorrect answers. The ninth question concerning the definition of a free radical had 76% correct answers, while 20% of the students chose the answer that defined it as a molecule capable of repairing cell damage, and only one participant thought it was a totally harmless waste product of the cell. The tenth question asked if they had ever heard of antioxidants; an

affirmative answer was given by 89% of the students. This knowledge was further investigated by asking them to choose between several options for the correct definition of antioxidants. 79% of the students gave the correct answer: that antioxidants are found both inside the cell and that they can be obtained through food. 14% of the students answered that they are molecules exclusively produced by the cell and therefore could not be obtained through food.

In question twelve, 96% of the students claimed they were familiar with omega-3. However, when asked in question thirteen which group of macromolecules omega-3 belonged to, only 13 students (44%) correctly answered 'fats,' while 49% believed it was proteins, and 7% thought it was carbohydrates. The data is illustrated in the histogram (see Figure 4).

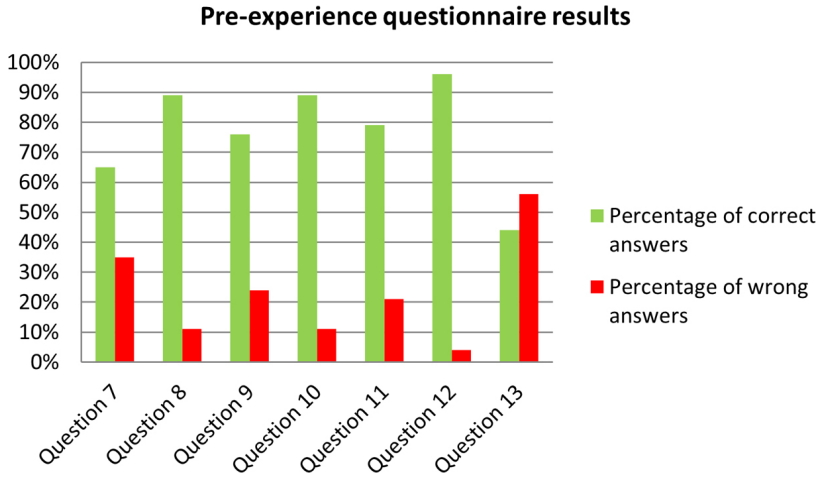

Fig. 4. Representative histogram of response rates to the pre-experience questionnaire

The post-experience questionnaire presented targeted questions about the topics shown in the game. The same questions from the pre-experience questionnaire were not repeated. Instead, an effort was made to use phrases very similar to those found in the explanatory panels of the virtual reality to evaluate how well they were read and understood. The first question reiterated the issue raised in the pre-experience questionnaire, which was the correct definition of a free radical. In this case, 90% of the responses were correct, with only three boys selecting an incorrect answer. This question already showed a more than 10% increase in correct responses. The percentage of correct answers regarding the definition of a free radical rose from 76% to 90%, representing an overall increase in 14% of correct responses. The second question inquired about the parameters that indicate the health status of a spermatozoon. 80% of the students provided the correct answer, while 17% considered an answer that also included the antioxidant content in the spermatozoon as correct. While this consideration may not be incorrect, since the qualitative parameters of the spermatozoon are explicitly displayed in the game and are criteria for progression, this answer was still deemed incorrect.

Only one student answered that the sperm parameters were tail length, vitality, and sperm potency. The third question, about what is correct to assume through food to counteract free radicals, received 97% correct answers. The correct answer was 'antioxidants'; the other answers were proteins, which received no selection, and fatty acids, which were selected by only one student. All in all, fatty acids are not a completely wrong answer, but considering the definition as generic and

not specifying what kind of fatty acids, the answer is to be considered wrong. The fourth question also had a high percentage of correct answers (97%). It asked how an 'antioxidant' could be defined. Only one student stated that antioxidants are nitrosamines, heterocyclic amines, and ethanol. A question concerning the definition of an antioxidant had already been proposed in the pre-experience questionnaire and had received a 79% correct answer rate. There was therefore an increase of 18% in correct answers. The fifth question, perhaps the most ambiguous of the entire post-experience questionnaire, asked why fatty acids are important for sperm health. 55% answered that it is because they constitute most of the spermatozoon's plasma membrane, which is the correct answer, but 41% answered that they counteract the action of free radicals. In the game, this issue is raised, but we see that it is the omega-3 polyunsaturated fatty acids (PUFA) and not generic fatty acids, that counteract free radical damage. These latter play a role in membrane repair. The next question, the sixth, was supposed to clarify if the concept of omega-3 as the most important fatty acid for sperm health was understood. It asked which of the proposed fatty acids were the most important for sperm health, and 93% of the students answered omega-3 and omega-6 PUFA. Only two students answered saturated fatty acids. At last, when asked which were the main food sources of omega-3 PUFA, 100% of the students gave the right answer, that is, oily fish, dried fruit, and green leafy vegetables. Compared to the pre-experience questionnaire, where 49% of the students had answered that omega-3 belonged to proteins, in this question it can be observed that following the VR experience, the students reconnected the source of omega-3 to what they had been shown in the game, therefore not as 'proteins' but as components of foods consisting of even more macromolecules, then implicitly identifying omega-3 among the fatty acids in the sixth question.

Data are shown in the histogram (see Figure 5).

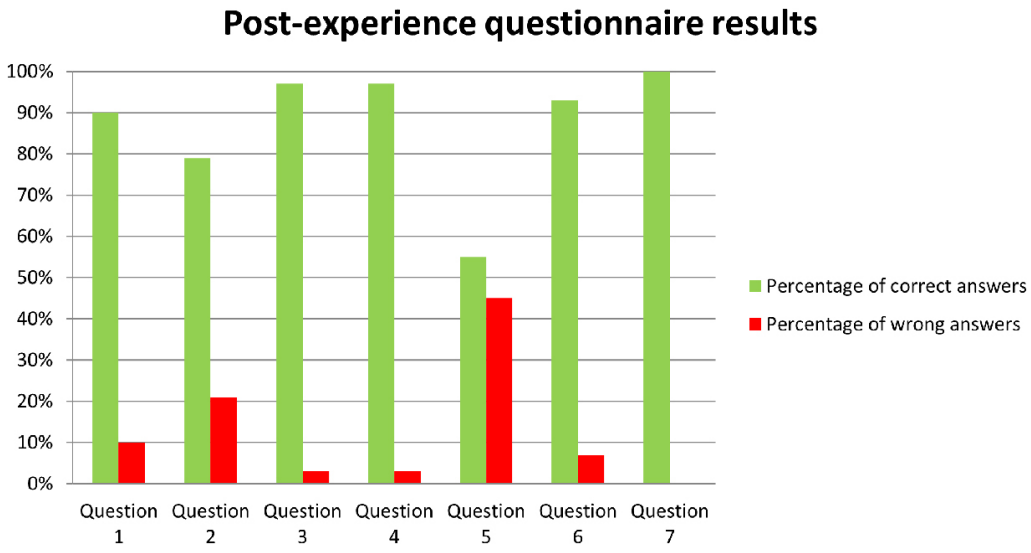

Fig. 5. Representative histogram representing the response rates to the post-experience questionnaire

When the results of the pre-experience questionnaire are compared with those obtained in the post-experience questionnaire, correct answers were achieved in at least 90% of the questions, with the exception of one question that may have been more ambiguous than the others.

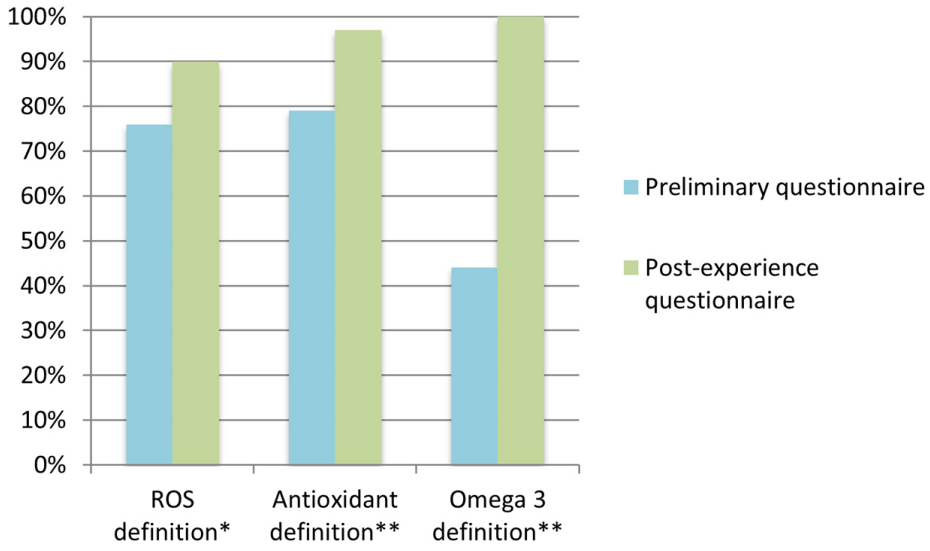

Fig. 6. Histogram comparing the correct answers with the same question topic in the pre-experience questionnaire versus the post-experience questionnaire

Note: An increase of at least 10% of the correct answers given after the virtual reality experience can be appreciated. $P < 0.05$ was considered significant. *$P < 0.05$; **$P < 0.01$.

When comparing questions with similar topics in the pre- and post-experience questionnaire, the number of correct answers significantly increased (ROS definition, $p < 0.05$, antioxidant and omega 3 definition $p < 0.01$ (see Figure 6)). The data was compared using the z-test for the binomial proportion directly performed on Excel.

4 DISCUSSION

Male fertility is declining. Studies have shown that dietary and lifestyle habits are crucial determinants of sperm function and fertility. Therefore, it is important to disseminate science-based knowledge on how a healthy diet can prevent certain male infertility [14]. A comprehensive systematic review of observational studies [9] revealed that omega-3 fatty acids, antioxidants such as vitamin E, vitamin C, β-carotene, selenium, zinc, cryptoxanthin, lycopene, vitamin D, folate, and low levels of saturated fatty acids and trans-fatty acids were inversely associated with low semen quality. Fish, shellfish, seafood, poultry, cereals, vegetables, and fruits were positively correlated with sperm quality parameters. Negative influences were also observed for processed meat, soy foods, potatoes, high alcohol intake, and caffeine. Animal models could help identify dietary therapies for reproductive diseases [15]. Men's fertility is significantly affected by obesity, diabetes, varicocele, and infections characterized by chronic inflammation, oxidative stress, and insulin resistance. In these pathological conditions, an increased level of ROS is counterbalanced by antioxidant treatment [16].

Dietary fatty acid supplementation influences the modulation of sperm mitochondria energy metabolism and, in turn, determines the improvement of sperm quality in rats [17]. These data underscore the importance of providing information on proper nutrition as a lifestyle choice, especially to young people in primary and secondary schools. In this context, a virtual game was developed where making correct food choices supports game progression. This game was tested on high school students. The playful aspect of the simulation is combined with an educational component that educates players about which foods in their daily diet contain antioxidants and omega-3, enabling users to connect the dynamics of the game to real life [13]. Two questionnaires

were administered: the pre-experience questionnaire was designed to evaluate the children's knowledge of nutrition. The responses we received indicated that a significant percentage of students are mindful of their daily food choices. Moreover, 25 out of 30 students believed it was important to rate the significance of food for their overall health between eight and 10. This suggests that the students participating in the study had a good understanding of the role that daily nutrition plays in their well-being.

A question on the definition of the Mediterranean diet was voluntarily included; a good percentage of students reported that the Mediterranean diet would involve a reduction of bread and pasta. This could be explained by the demonization of carbohydrates by certain strict dietary protocols that have become increasingly popular on social media, such as the ketogenic diet. However, the scientifically defined Mediterranean diet, and the only diet approved even in the long term along with vegetarian diets for sustaining optimal health, includes a predominant presence in the daily diet of carbohydrates, derived from whole grains or from grains themselves in grain form, up to 55%–60% of the macronutrient composition of the day's diet. This data suggests that in food prevention for both primary and secondary school students, 'playful' components such as our experience in VR can be exploited to properly nourish our organism and to have a tool for understanding and filtering the most disparate information that is now readily available on social media and the Internet [18, 19]. Overall, the questions asked to understand the degree of biological knowledge and interest in food revealed that there was a good awareness of the importance of nutrition, and some basic information seemed clear. This could be justified by the fact that the sample of students we selected came from a school focused on education in science subjects. This suggests that to support the result of this study, a wider range of participants should be included in the future study, opening it up to schools with humanities subjects. 65% of these students had already studied the spermatozoon, and 89% recognized the correct definition of this cell. Free radicals were correctly defined by 76% of the students. 90% of them said they had already heard of antioxidants, but the correct definition was given by only 79% of them. The same situation was found for omega-3; 96% said they were already familiar with them, but 48% of the students defined them as 'proteins.' This makes one reflect on the kind of marketing that has been done in recent years in the field of food, which has seen protein described as the main macronutrient for the health and well-being of the human body, when scientifically an excess of protein is not recommended, with 0.8–1.0 g/kg/day [20] being sufficient for a healthy adult, and an excess can be deleterious in certain diseases such as chronic renal failure or liver failure. Two levels of dietary attention are presented in the game, the first showing how a diet rich in antioxidants provides the player with a reserve of munitions to scavenge ROS in seminal fluid and thus improves seminal motility.

The study demonstrates how dietary choices that prioritize omega-3 intake can help repair damage to the spermatozoa's plasma membrane. In the post-experiment questionnaire, students indicated that they had acquired specific knowledge related to their virtual reality experience. With correct answers exceeding 90% in almost all questions (except for one that was slightly more ambiguous), it can be concluded that the participants in this experiment, despite their initial good knowledge base and the necessity to broaden the study's scope, have enhanced their understanding of nutrition and reproductive biology.

The game allowed participants to expand their knowledge of diet and consolidate their previous statements. When asked what a free radical was in the pre-experiment questionnaire, the percentage of correct answers was around 70%. In the post-experiment questionnaire, the percentage increased to 85%. Students have

shown the greatest gaps in the definition and role of fatty acids in sperm health. Fatty acids are crucial for sperm health as they constitute the majority of the spermatozoa's plasma membrane, as is the case in all eukaryotic cells. This discovery suggests that the sperm cell serves as a model that can be extrapolated to all cells. The function of fatty acids is not to combat free radicals but to repair cell membranes previously damaged by them. Nonetheless, 93% of the students identified omega-3 and omega-6 PUFA as essential for sperm quality. Based on the study's findings, it can be inferred that utilizing games, particularly virtual reality, is a significant method to enhance the understanding of complex topics, ensuring the transmission of information through images and experiences endures over time. In a previous study, assessments of sickness, presence, and usability confirmed that the game was user-friendly and safe to use [21]. The study's limitations, as previously mentioned, include the small sample size of 30 students, all of whom were enrolled in a school with technical-scientific subjects. For future studies, it would be advisable to broaden the study population both quantitatively and in terms of foundational knowledge, involving institutions and high schools with diverse subject areas, to determine if virtual reality can indeed help students from non-scientific backgrounds comprehend unfamiliar topics.

Other limitations now concern the spread and cost of virtual reality; however, we must consider that the purchase of a visor in recent months has reached the cost of a normal video game, and the price will fall further, allowing students to enjoy the learning experience even at a distance. A subsequent experiment could be to compare two homogeneous groups of students who acquired knowledge in different ways: one group through traditional study (a textbook) versus another who took advantage of virtual reality. It would then be interesting, several months or a few years later, to test again using the post-experience questionnaire to see how much of what the students experienced in the virtual reality experience stayed in their memory. This assessment could be used to compare with the academic study to determine how much information lasts over time thanks to an immersive and personifying experience such as virtual reality. Combining play with learning healthy behaviors could be a new opportunity to access otherwise complex information in a simple way, to unconsciously change one's lifestyle, and to significantly prevent various diseases or disabling situations, such as infertility.

5 REFERENCES

[1] L. Wiewiorra and I. Godlovitch, "The digital services act and the digital markets act – a forward-looking and consumer-centred perspective," *Publication for the Committee on the Internal Market and Consumer Protection, Policy Department for Economic, Scientific and Quality of Life Policies*, European Parliament, Luxembourg, 2021.

[2] M. Slater and M. V. Sanchez-Vives, "Enhancing our lives with immersive virtual reality," *Front. Robot. and AI*, vol. 3, 2016. https://doi.org/10.3389/frobt.2016.00074

[3] B. Ferguson, "The emergence of games for health," *Games Health J.*, vol. 1, no. 1, pp. 1–2, 2012. https://doi.org/10.1089/g4h.2012.1010

[4] J. Fox and J. N. Bailenson, "Virtual self–modeling: The effects of vicarious reinforcement and identification on exercise behaviors," *Media Psychol.,* vol. 12, no. 1, pp. 1–25, 2009. https://doi.org/10.1080/15213260802669474

[5] K. P. Phillips and N. Tanphaichitr, "Human exposure to endocrine disrupters and semen quality," *J. Toxicol. Environ. Health, Part B,* vol. 11, nos. 3–4, pp. 188–220, 2008. https://doi.org/10.1080/10937400701873472

[6] C. H. Ramlau-Hansen, A. M. Thulstrup, A. S. Aggerholm, M. S. Jensen, G. Toft, and J. P. Bonde, "Is smoking a risk factor for decreased semen quality? A cross-sectional analysis," *Hum. Reprod.,* vol. 22, no. 1, pp. 188–196, 2007. https://doi.org/10.1093/humrep/del364

[7] R. Finelli, F. Mottola, and A. Agarwal, "Impact of alcohol consumption on male fertility potential: A narrative review," *Int. J. Environ. Res. Public Health*, vol. 19, no. 1, p. 328, 2021. https://doi.org/10.3390/ijerph19010328

[8] H. Levine *et al.*, "Temporal trends in sperm count: A systematic review and meta-regression analysis of samples collected globally in the 20th and 21st centuries," *Hum. Reprod. Update*, vol. 29, no. 2, pp. 157–176, 2023. https://doi.org/10.1093/humupd/dmac035

[9] A. Salas-Huetos, M. Bulló, and J. Salas-Salvadó, "Dietary patterns, foods and nutrients in male fertility parameters and fecundability: A systematic review of observational studies," *Hum. Reprod. Update*., vol. 23, no. 4, pp. 371–389, 2017. https://doi.org/10.1093/humupd/dmx006

[10] A. Ferramosca and V. Zara. "Diet and male fertility: The impact of nutrients and antioxidants on sperm energetic metabolism," *Int. J. Mol. Sci*., vol. 23, no. 5, p. 2542, 2022. https://doi.org/10.3390/ijms23052542

[11] S. Alesi *et al.*, "Anti-inflammatory diets in fertility: An evidence review," *Nutrients*, vol. 14, no. 19, p. 3914, 2022. https://doi.org/10.3390/nu14193914

[12] C. Castellini *et al.*, "Effect of dietary *n*-3 Source on rabbit male reproduction," *Oxid. Med. Cell. Longev*., vol. 2019, no. 1, p. 3279670, 2019. https://doi.org/10.1155/2019/3279670

[13] G. Collodel *et al.*, "Antioxidants, dietary fatty acids, and sperm: A virtual reality applied game for scientific dissemination," *Oxid. Med. Cell. Longev*., vol. 2019, no. 1, p. 2065639, 2019. https://doi.org/10.1155/2019/2065639

[14] A. J. Gaskins and J. E. Chavarro, "Diet and fertility: A review," *Am. J. Obstet. Gynecol.*, vol. 218, no. 4, pp. 379–389, 2018. https://doi.org/10.1016/j.ajog.2017.08.010

[15] F. L. Nassan, J. E. Chavarro, and C. Tanrikut, "Diet and men's fertility: Does diet affect sperm quality?" *Fertil. Steril*., vol. 110, no. 4, pp. 570–577, 2018. https://doi.org/10.1016/j.fertnstert.2018.05.025

[16] A. Ferramosca, N. Moscatelli, M. Di Giacomo, and V. Zara, "Dietary fatty acids influence sperm quality and function," *Andrology*, vol. 5, no. 3, pp. 423–430, 2017. https://doi.org/10.1111/andr.12348

[17] J. Whittaker, "Dietary trends and the decline in male reproductive health," *Hormones*, vol. 22, pp. 165–197, 2023. https://doi.org/10.1007/s42000-023-00431-z

[18] K. Rounsefell *et al.*, "Social media, body image and food choices in healthy young adults: A mixed methods systematic review," *Nutr. Diet*., vol. 77, no. 1, pp. 19–40, 2020. https://doi.org/10.1111/1747-0080.12581

[19] A. Chung *et al.*, "Adolescent peer influence on eating behaviors via social media: Scoping review," *J. Med. Internet Res*., vol. 23, no. 6, p. e19697, 2021. https://doi.org/10.2196/19697

[20] SINU, "Fabbisogno proteico" and "Livelli di assunzione di riferimento per le proteine e adeguatezza nella dieta," in *Livelli di Assunzione di Riferimento di Nutrienti ed Energia per la popolazione italiana, IV revision*, Eds., Milano: SICS editore Srl, 2014, pp. 161–172.

[21] S. Ermini, G. Collodel, A. Innocenti, M. Masini, E. Moretti, and V. Santalucia, "Immersive virtual reality as a tool for education: A case study," in *Handbook of Research on Implementing Digital Reality and Interactive Technologies to Achieve Society 5.0*, F. M. Ugliotti and A. Osello, Eds., IGI Global, 2022, pp. 121–136. https://doi.org/10.4018/978-1-6684-4854-0.ch006

6 AUTHORS

Elena Bolognesi was with the Department of Molecular and Developmental Medicine, University of Siena, Policlinico Le Scotte, Viale Bracci, 14, 53100 Siena, Italy. She is now a freelance dietician.

Elena Moretti is with the Department of Molecular and Developmental Medicine, University of Siena, Policlinico Le Scotte, Viale Bracci, 14, 53100 Siena, Italy.

Maurizio Masini is with the Department of Social, Political and Cognitive Sciences, University of Siena, Palazzo San Niccolò, via Roma, 56, 53100 Siena, Italy.

Cinzia Signorini is with the Department of Molecular and Developmental Medicine, University of Siena, Policlinico Le Scotte, Viale Bracci, 14, 53100 Siena, Italy.

Vincenzo Santalucia is with the Department of Social, Political and Cognitive Sciences, University of Siena, Palazzo San Niccolò, via Roma, 56, 53100 Siena, Italy.

Giulia Collodel is with the Department of Molecular and Developmental Medicine, University of Siena, Policlinico Le Scotte, Viale Bracci, 14, 53100 Siena, Italy. She is the corresponding author, E-mail: giulia.collodel@unisi.it.

Online-Journals.org

iJET International Journal of Emerging Technologies in Learning

iJET | eISSN: 1863-0383 | Vol. 19 No. 6 (2024) | OPEN ACCESS

https://doi.org/10.3991/ijet.v19i06.49799

PAPER

The Effect of ChatGPT on Education in the UAE

Mohammed Alnahhal[1](✉), Hamda Alali[2], Reema Alshamsi[3]

[1]Mechanical and Industrial Engineering Department, American University of Ras Al Khaimah, Ras Al Khaimah, United Arab Emirates

[2]Department of Computer Science and Engineering, American University of Ras Al Khaimah, Ras Al Khaimah, United Arab Emirates

[3]Chemical and Petroleum Engineering Department, American University of Ras Al Khaimah, Ras Al Khaimah, United Arab Emirates

mohammed.alnahhal@aurak.ac.ae

ABSTRACT

This study examines the impact of ChatGPT on undergraduate students' learning. The study utilizes a survey to gather authentic data from a sample of 316 bachelor students from 15 universities across the UAE. The survey comprises 22 dimensions categorized into four constructs related to ChatGPT usage, benefits, drawbacks, and the types of questions ChatGPT can address. A statistical analysis was conducted to explore three primary hypotheses. The analysis involved using Excel and R software to perform one-sample t-tests, two-sample t-tests, ANOVA, and correlation. The findings indicate that students' use of ChatGPT for various assignments is still infrequent, despite their familiarity with it. This suggests that educators have not yet integrated ChatGPT into assignments and view it more as a threat than an opportunity. Students acknowledge both the advantages and disadvantages of ChatGPT. However, students in medical fields expressed more criticism towards ChatGPT compared to students in engineering, computer science, and business disciplines. Additionally, male and female students hold similar perceptions, except for two dimensions. The results also suggest that educators should pose questions that necessitate personal arguments, experiences, and critical thinking to deter academic dishonesty.

KEYWORDS

ChatGPT, educators, learning, cheating, innovative assignments

1 INTRODUCTION

Digitalization has emerged very rapidly in the last few years and is influencing human life, including education. Big data is a major outcome of digitization, and it is an area of investment for many organizations contemplating knowledge mining and extracting relevant knowledge. ChatGPT is a recent example of an innovation utilizing artificial intelligence (AI) that applies machine learning to extract useful knowledge from large amounts of data. ChatGPT was introduced in late 2022 by OpenAI, and while it is not the first AI model ever created nor the first by OpenAI, it marks a significant advancement in the field [1]. ChatGPT is a large model to perform natural language processing (NLP) tasks using deep learning from large amounts of data. Students in universities have started to utilize ChatGPT in their assignments,

Alnahhal, M., Alali, H., Alshamsi, R. (2024). The Effect of ChatGPT on Education in the UAE. *International Journal of Emerging Technologies in Learning (iJET)*, 19(6), pp. 65–78. https://doi.org/10.3991/ijet.v19i06.49799

Article submitted 2024-04-23. Revision uploaded 2024-06-01. Final acceptance 2024-06-02.

and researchers are currently using it to prepare their study papers. This has raised some concerns about the use of ChatGPT in education, such as generating false data and plagiarism [2, 3].

The aim of this paper is to investigate the positive and negative effects of ChatGPT on bachelor students' education based on their perceptions. Since the ChatGPT concept is new, study is still needed to learn about its advantages and disadvantages. This is done using real data from the users themselves. Therefore, a survey is used to gather the perceptions of 316 students in 15 universities from various majors in the UAE about ChatGPT opportunities, threats, and potential ways to use it effectively. To the best of the authors' knowledge, no previous studies have conducted such a comprehensive analysis in the UAE. Ammar et al. [4] investigated the effect of ChatGPT on student behavior in the UAE, but on a small sample size of 54 participants.

The use of ChatGPT presents challenges in education. Both students and professors need training on how to use it efficiently [5]. In ChatGPT, the dialog setup allows users to repeatedly change their input to get clearer and more accurate answers. However, the model's output result cannot be guaranteed to be accurate and must be cross-checked, taking context and knowledge domains into account. Based on complex math and probability science, ChatGPT can uncover patterns and extract relationships between words and topics. But what about context, meaning, text thoughts, and reasoning? ChatGPT doesn't take any responsibility for the output of the model. If the output of the model is input to another system, who will be responsible for the consequences? ChatGPT and other similar tools are still incompetent and far from intelligent [6]. ChatGPT sparked a discussion about the potential benefits and threats, particularly for higher education institutions. Many researchers discussed the opportunity of enhancing learning for both students and faculty. In addition, ChatGPT may help researchers develop their study work, considering academic integrity. Other researchers have spoken about academic integrity risks [1, 7]. Threats include issues with work originality, intellectual property, honesty, integrity, and authorship, all of which are crucial academic principles.

2 LITERATURE REVIEW

Artificial intelligence (AI) is rapidly evolving and increasingly influencing our daily lives. The ability to process and learn from various types of data to extract useful and structured knowledge for automating human tasks or supporting detection holds a promising future [8]. The advancements in AI technologies, supported by the rapid development of deep machine learning, have led to the creation of large processing models like ChatGPT. ChatGPT is a language model trained on extensive datasets, including books, articles, websites, etc. [9]. This model can extract structured text and generate coherent and meaningful text using advanced deep-learning algorithms for NLP tasks. The learning process aims to identify patterns, trends, and relationships in texts to produce fully structured topics across various domains of knowledge [10].

ChatGPT benefits are acknowledged by many researchers, despite some disagreement. The ChatGPT model could assist in solving complex problems for learners [10]. Furthermore, its dialog capability and ability to respond to human natural language queries increase its appeal for learning purposes [11]. ChatGPT can also help learners grasp concepts from various knowledge domains before delving into formal educational resources such as books and lectures. Additionally, it can enhance

students' writing skills by assisting in text composition or restructuring sentences in alternative formats [12]. General language models such as ChatGPT are valuable for summarizing large bodies of text and abstracting information. This feature is particularly beneficial for students and researchers seeking to quickly gather the necessary knowledge for learning and writing study reports [13, 14]. ChatGPT is considered a valuable tool for both students and teachers in facilitating the learning process. For students, ChatGPT aids in self-study and extracting diverse and engaging learning materials such as text, images, and videos [15].

Scientific researchers may also use ChatGPT mainly for editing and paraphrasing. ChatGPT can provide alternative words to improve sentence structure for better clarity [12]. According to Hutson [16], some scientific researchers utilize AI-based tools to aid in evaluating, coding, and summarizing literature reviews. Researchers have identified and listed six core values of academic integrity: honesty, trust, fairness, respect, responsibility, and courage. Using ChatGPT to create reports, research, or essays violates academic integrity. Additionally, students or researchers can engage with ChatGPT to receive various outputs and select the most suitable for their work. The output from ChatGPT is generated from a vast collection of others' original work, raising significant concerns about intellectual property rights [1]. The accuracy of ChatGPT is another critical concern that questions the credibility of its outputs. The accuracy of ChatGPT is influenced by the volume of data it has been trained on, the complexity of the query, and the domain knowledge. However, it is highly recommended that the outputs be double-checked and validated, considering the context and domain knowledge [17].

Michalak [18] suggests that authors will soon be required to declare whether they used an AI-generated model to produce their study articles. Sullivan, Kelly, and McLaughlan [19] suggest that universities and higher education institutions should examine and explore the opportunity of using AI tools such as ChatGPT to enhance student learning. Utilizing this tool could benefit any type of learner at any level. However, like any other high-technology tool in our current era, it can be misused. Educational institutions and universities need to understand the tools' capabilities, accuracy, credibility, and scalability limitations and then develop policies, guidelines, and procedures for how to use any AI language model, such as ChatGPT.

Researchers investigated the impact of ChatGPT on education through various approaches, often focusing on specific fields or students from particular majors. For instance, Rejeb et al. [3] employed machine learning models to analyze 2003 web articles, highlighting the benefits and concerns regarding ChatGPT in education, including ethical considerations like data privacy. Halaweh [20] explored strategies for the responsible and effective integration of ChatGPT in education, presenting five techniques for student use. Srinivasan [21] envisioned a promising future for AI in education, offering a comprehensive framework to maximize learning potential and positive outcomes. Wardat et al. [17] studied the impact of ChatGPT on teaching and learning mathematics, while Ali et al. [22] examined its influence on English learning motivation. Firat [23] surveyed scholars and students from various countries to analyze perceptions of the advantages and challenges of ChatGPT. Al Shloul et al. [24] used a comparative approach to demonstrate the benefits of ChatGPT in education, such as creating an interactive learning environment. Chaudhry et al. [25] employed an experimental design to assess ChatGPT's ability to solve diverse assignments. Rasul et al. [26] identified five benefits and challenges of ChatGPT in education through a literature review. Student perceptions of ChatGPT were evaluated through a survey in [27], focusing solely on senior students in a

computer engineering program. In contrast, our study encompasses a large sample of students from 15 universities and various majors, providing a comprehensive assessment of these students' perceptions.

3 METHODOLOGY

The questionnaire in this study was distributed to many students in April 2023. The respondents are mainly bachelor students from 15 universities across the UAE, including the top-ranked universities in the country such as Sharjah University, United Arab Emirates University, Khalifa University, Abu Dhabi University, Ajman University, Higher Colleges of Technology, and Zayed University. However, the largest percentage of respondents are from the American University of Ras Al Khaimah. The total number of participants is 316, making the sample size large enough to represent the population. The study is based on the analysis of a questionnaire that focuses on four constructs: ChatGPT usage, ChatGPT advantages, ChatGPT disadvantages, and ChatGPT usage management. Each construct contains dimensions for measurement, totaling 22 dimensions. These dimensions were designed based on a literature review and expert opinions. In addition to these questions, there are three questions at the beginning of the questionnaire about the gender, specialty, and class standing levels (freshman, sophomore, junior, senior) or study year (1, 2, 3, or 4).

Table 1 presents the constructs and dimensions of the study. The questionnaire comprises questions related to the dimensions in Table 1, along with basic demographic data such as gender, academic level, and major or field of specialization. This concise overview is provided for future reference in other tables. The questionnaire consists of a total of 25 questions. Respondents are required to select one of the following options for the 22 dimensions: strongly agree, agree, not sure, disagree, or strongly disagree. During the analysis, the responses are transformed into a Likert scale ranging from 1 to 5, with 5 indicating strongly agree and 1 indicating strongly disagree.

Table 1. Study constructs and dimensions

Constructs	Dimension	Short Expression
ChatGPT Usage	I am familiar with ChatGPT	familiar with ChatGPT
	I use ChatGPT to answer homework questions	to answer homework
	I used ChatGPT for major assignments or projects	for major assignments
	I use AI tools other than ChatGPT	other tools
	I do NOT face technical troubles when I use ChatGPT	no technical troubles
ChatGPT Advantages	ChatGPT advantages are more than disadvantages	advantages are more
	I have a positive experience with ChatGPT	positive experience
	ChatGPT understands and answers the questions of the students correctly	understands and answers
	ChatGPT is capable of answering different types of questions	answering different questions
	I intend to use ChatGPT in my future career	use in future
	ChatGPT can improve student learning process	improve learning

(Continued)

Table 1. Study constructs and dimensions *(Continued)*

Constructs	Dimension	Short Expression
ChatGPT Disadvantages	ChatGPT can be misused	misused
	ChatGPT can be used in cheating in academia	cheating
	ChatGPT can destroy a lot of jobs in the future	destroy jobs
	ChatGPT makes students lazy and not making their jobs by themselves	lazy students
	ChatGPT still needs to be enhanced to work better in helping students	enhancements needed
Managing using ChatGPT	ChatGPT is always correct in answering calculations questions	correct calculations
	ChatGPT is a learning aid and it represents a huge opportunity for learning	learning aid
	Educators should introduce innovative assessments to use ChatGPT	innovative assessments needed
	ChatGPT can be used by students even when they are asked to include personal experiences or perspectives in their writing	effect of personal experiences
	ChatGPT can be used by students even when they are asked to include their own argument about a subject	effect of own argument
	ChatGPT can be used even for assignments that need creative and critical thinking abilities	effect of critical thinking

Dividing the study measures into constructs and dimensions makes the analysis easier to interpret. However, tests should be conducted to assess the internal consistency of each construct. Cronbach's alpha is used for this purpose. The impact of gender, major, and academic level will be examined. This examination is structured around the study hypotheses, which are outlined below:

- H1: Students agree or strongly agree about ChatGPT usage, advantages, disadvantages, and management.
- H2: The respondents' perception of ChatGPT may be influenced by gender.
- H3: The respondents' perception of ChatGPT may be influenced by factors such as class standing and major.

The above hypotheses are the "alternative" ones. The "null" hypotheses mean that there is no effect of gender, level, or major. Rejecting the null hypothesis means that the alternative hypothesis is accepted. Initially, basic statistics such as average and standard deviation are calculated for each question. The first hypothesis is investigated using a one-sample t-test. The mean will be checked if it exceeds 3 ($\mu > 3$) because exceeding 3 indicates "agree" (4 on the Likert scale) or "strongly agree" (5 on the Likert scale). Rejecting the null hypothesis implies that the gender of the students affects their perception. To conduct a t-test, the p-value is calculated, and if the p-value is less than 0.05, the null hypothesis is rejected. This principle regarding the p-value applies to all three hypotheses. For the first hypothesis, if the average is less than 3, a t-test is unnecessary because the average cannot be greater than 3. The second hypothesis is examined using a two-sample t-test, assuming unequal variances. This test is typically conducted when there

are two samples, such as a sample of males and a sample of females. This process is carried out for each of the 22 questions. A two-sided t-test is employed, which means that the alternative hypotheses check if $\mu 1 \neq \mu 2$. In the third hypothesis, there are more than two samples, for instance, four levels and five major groups. If at least one of the levels differs from the others, the null hypothesis is rejected. For instance, if senior students have different perceptions than other students, the alternative hypothesis is accepted. This procedure is also applied to each question for the two factors (major and level). In this scenario, a t-test is unnecessary; instead, Analysis of Variance (ANOVA) is utilized. Consequently, three tests are conducted in this study, each on every question. To analyze the construct, the average of the respondents' answers about the dimensions of each construct is utilized.

Besides the above hypotheses, it is expected that individuals who use ChatGPT more frequently will hold more positive perspectives regarding its advantages. Therefore, the correlation between the frequency of usage and advantages is anticipated to be stronger than the correlation between usage and disadvantages. The Spearman correlation coefficient is calculated to verify this. The entire analysis is conducted using Excel and R software.

4 ANALYSIS AND RESULTS

The number of male and female respondents is shown in Figure 1. The number of males and females is almost the same. Figure 2 shows the majors of the respondents. The number of fields of specialization is 58, which are combined into major groups. A large percentage of students are in the category of computer-related fields such as IT, AI, computer engineering, computer science, cybersecurity, and others. The second category comprises the rest of engineering majors such as mechanical engineering, electrical engineering, industrial engineering, etc. Additionally, medicine-related fields include medicine, dentistry, human nutrition and dietetics, physical therapy, nursing, and others. Business majors consist of business administration, business analytics, business finance, business marketing, digital marketing, and others. Subsequently, in the analysis, those majors with fewer than 10 students will be combined and named as "others". Figure 3 shows the students' class standing. Students are from all levels.

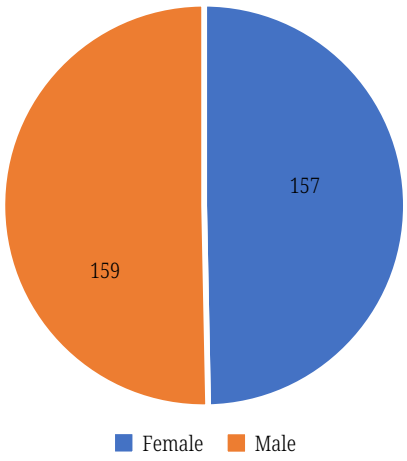

Fig. 1. Gender of respondents

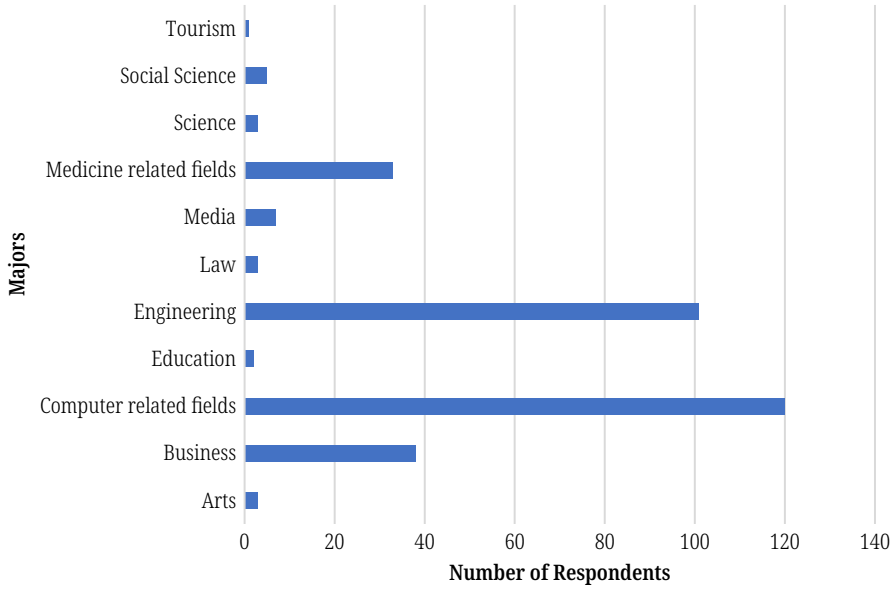

Fig. 2. Majors of respondents

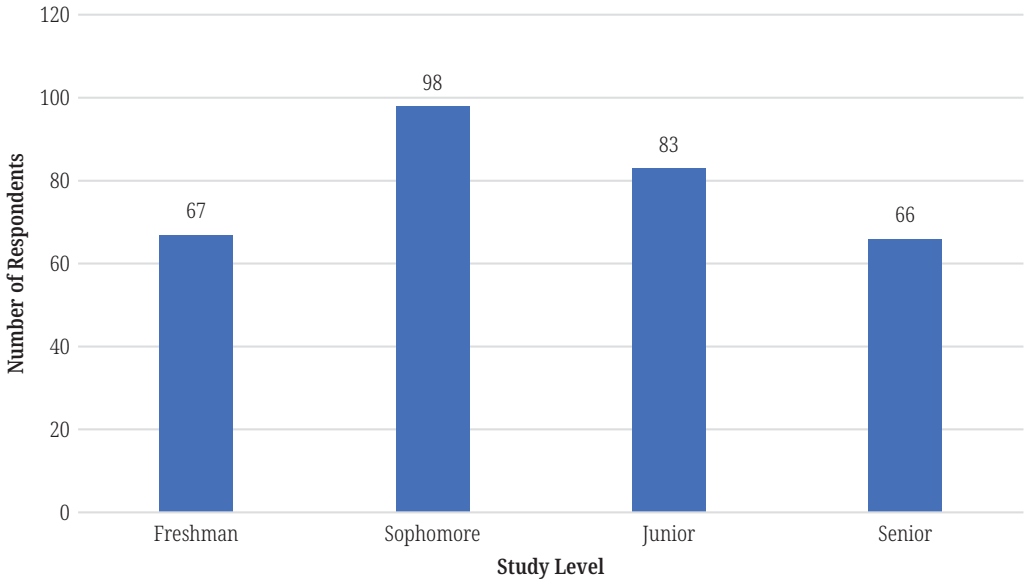

Fig. 3. Respondents class standing levels

The internal consistency of the dimensions in each construct is assessed using Cronbach's alpha. Table 2 displays these values. All values are sufficiently high, as they are all above 0.6. The table also indicates that the overall average for each construct is approximately 3 or higher.

Table 2. Study constructs and averages

Constructs	Average	Cronbach Alpha
ChatGPT Usage	3.01	0.68
ChatGPT Advantages	3.20	0.82
ChatGPT Disadvantages	3.27	0.75
Managing using ChatGPT	3.19	0.67

Table 3 presents the averages, standard deviations, and t-test results for each dimension (question). NA is used when the average is less than 3. It is evident that respondents generally agree on the advantages, disadvantages, and management of ChatGPT. However, ChatGPT is not extensively utilized by students. Although students are acquainted with ChatGPT, they do not use it regularly. This is apparent as the averages are below 3 in four questions of the first construct, eliminating the need for a t-test.

Table 3. Average, standard deviation, and one sample t-test

	Average	Standard Deviation	t-Test (p-Value)
ChatGPT Usage			0.42
familiar with ChatGPT	3.45	1.35	0.00
to answer homework	2.92	1.26	NA
for major assignments	2.92	1.27	NA
other tools	2.81	1.32	NA
no technical troubles	2.94	1.16	NA
ChatGPT Advantages			0.00
advantages are more	3.19	1.26	0.00
positive experience	3.27	1.27	0.00
understands and answers	3.15	1.15	0.01
answering different questions	3.31	1.20	0.00
use in future	3.22	1.20	0.00
improve learning	3.08	1.26	0.14
ChatGPT Disadvantages			0.00
misused	3.31	1.28	0.00
cheating	3.41	1.32	0.00
destroy jobs	3.30	1.24	0.00
lazy students	3.10	1.28	0.09
enhancements needed	3.23	1.20	0.00
Managing using ChatGPT			0.00
correct calculations	3.39	1.21	0.00
learning aid	3.27	1.17	0.00
innovative assessments needed	3.26	1.14	0.00
effect of personal experiences	3.14	1.24	0.03
effect of own argument	3.04	1.19	0.27
effect of critical thinking	3.07	1.19	0.15

The students agreed about almost all the advantages mentioned. This can be seen when the p-value is less than 0.05. However, many of them do not think it can improve learning (p-value = 0.14). The students are concerned about the disadvantages of

ChatGPT, where the p-values are zero in most of them. The management of using ChatGPT is needed. Depending on the p-values, we can say that the students think that ChatGPT cannot be useful for questions that require their own argument and critical thinking. However, it can be very useful for calculation questions. For the questions about personal experiences, males and females have different perspectives. In the next paragraph, this will be discussed.

The above explanation examines the first hypothesis. For the second hypothesis, a two-sample t-test is required to assess the impact of gender on the results. The p-values are greater than 0.05 in all questions except for questions 15 and 20. Figure 4 illustrates the contrasting viewpoints of males and females on these two questions. Females believe that ChatGPT can remain beneficial even when questions about personal experience are required. Conversely, males believe that ChatGPT can lead to student laziness.

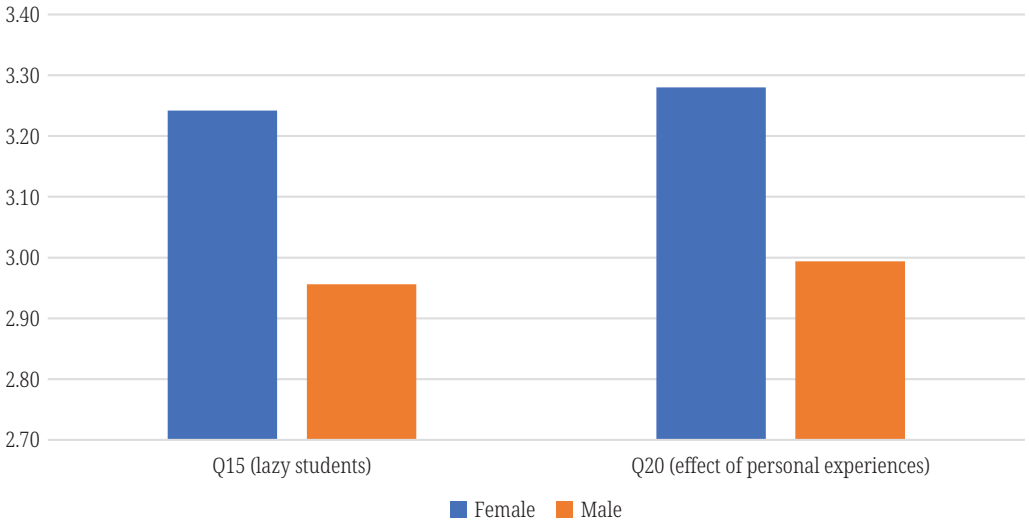

Fig. 4. Effect of gender on the perception of respondents regarding Q15 and Q20

The third hypothesis examines the impact of class standing level and major on students' perceptions. An ANOVA test was conducted, revealing that students across different class standing levels share similar perceptions of ChatGPT, as indicated by p-values exceeding 0.05. However, the influence of major is more pronounced. By excluding majors with a small number of students, the sample size is reduced to 292. Table 4 displays the results of the ANOVA test for the questions.

Table 4. ANOVA test to check the effect of specialty

Question #	Constructs and Dimensions (Short Expression)	ANOVA Test (P-Value)
	ChatGPT Usage	
Q1	familiar with ChatGPT	0.87
Q2	to answer homework	0.03
Q3	for major assignments	0.01
Q4	other tools	0.07
Q5	no technical troubles	0.24

(Continued)

Table 4. ANOVA test to check the effect of specialty *(Continued)*

Question #	Constructs and Dimensions (Short Expression)	ANOVA Test (P-Value)
	ChatGPT Advantages	
Q6	advantages are more	0.91
Q7	positive experience	0.99
Q8	understands and answers	0.28
Q9	answering different questions	0.07
Q10	use in future	0.23
Q11	improve learning	0.82
	ChatGPT Disadvantages	
Q12	misused	0.00
Q13	cheating	0.00
Q14	destroy jobs	0.03
Q15	lazy students	0.03
Q16	enhancements needed	0.32
	Managing using ChatGPT	
Q17	correct calculations	0.02
Q18	learning aid	0.13
Q19	innovative assessments needed	0.46
Q20	effect of personal experiences	0.11
Q21	effect of own argument	0.54
Q22	effect of critical thinking	0.23

Figure 5 shows the varying perceptions of students from disparate major groups. The differences in other questions, not depicted in Figure 5, were found to be insignificant, with p-values exceeding 0.05. Hence, they have not been included in the figure.

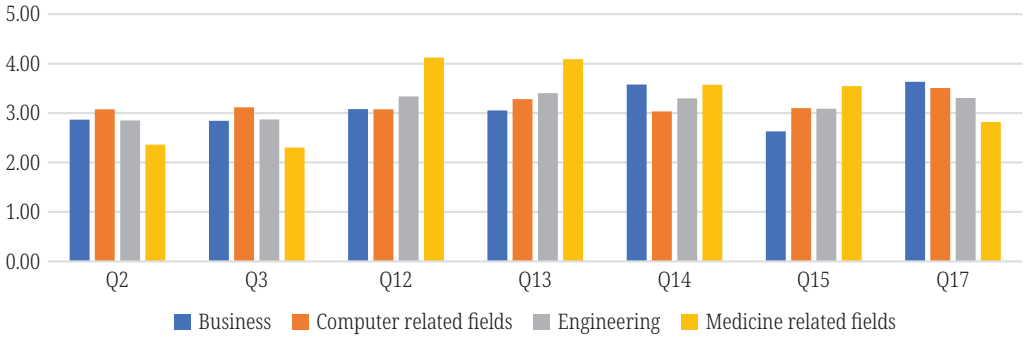

Fig. 5. Effect of specialty on the perception of respondents

In spite of the differences between questions Q2 and Q3, almost all groups of students rated the questions lower than 3. Only the students in the computer-related fields have an average slightly larger than 3. These two questions are about the use

of ChatGPT. For questions 12, 13, 14, and 15, which are about the disadvantages of ChatGPT, the students from the medicine-related fields have the largest criticism about ChatGPT. The correlation coefficient was used to check if the students who use ChatGPT are more convinced about its advantages. Table 5 shows the correlation between the usage of ChatGPT and the perception of its advantages and disadvantages. The correlation between usage and advantages is relatively high (0.56). However, the correlation between usage and disadvantages can be negligible.

Table 5. Correlation between ChatGPT usage and its advantages and disadvantages

	ChatGPT Advantages	ChatGPT Disadvantages
ChatGPT usage	0.56	0.19

The main results of the study indicate a lack of awareness regarding the full benefits of ChatGPT. Students are more aware of its disadvantages than its advantages. Part of the problem is that ChatGPT is still in its initial phase of usage. Typically, professors avoid creating assignments specifically designed to use ChatGPT. The different specialties of students affect their perceptions. The study results generally agree with the major findings of previous studies. This is especially relevant to the importance of designing assignments in a way to prohibit cheating [2, 28]. Generally, the study shows that students are familiar with ChatGPT but are not using it frequently. However, in the future, it is expected that ChatGPT will become more popular. Students believe in its advantages and disadvantages. However, students from medicine-related fields have a stronger perception of ChatGPT's disadvantages. The performance of ChatGPT depends on the subject domain. This is the same result found by Lo [5]. Moreover, ChatGPT can be used more for simple, direct questions such as calculations. However, for those questions that include opinions, own arguments, or critical thinking, ChatGPT cannot be useful. Educators should design questions in an innovative way to utilize ChatGPT to some extent and also encourage students to exert efforts to answer the questions of assignments. ChatGPT can be used to create engaging and challenging tasks. Based on the above, the following managerial implications can be highlighted:

- Universities should not avoid using ChatGPT due to the challenges it presents. Instead, they should carefully manage the way ChatGPT is used.
- Universities should provide training for both students and faculty members on how to utilize ChatGPT efficiently. This training should cover not only the technical aspects of using the tool but also ethical considerations and proper citation practices.
- Professors should design assignments in a way that prevents cheating, such as by requiring personal perspectives from students.

Overall, embracing ChatGPT in education requires a balanced approach that addresses its challenges while considering its potential benefits to enhance learning outcomes.

5 CONCLUSION

In this study, we investigated the use of ChatGPT by bachelor students at different universities in the UAE. Data was collected through a questionnaire that asked students about using ChatGPT to solve their assignments. It also asked them about

their opinions regarding the advantages and disadvantages of ChatGPT and the types of questions ChatGPT can answer. Statistical analyses such as the one-sample t-test, two-sample t-test, ANOVA, and correlation were used to investigate three main hypotheses. Results showed that ChatGPT usage in assignments is still in its early stages. Students are familiar with ChatGPT and its pros and cons. The major has a significant impact on results. For instance, students in the medical field are more critical of its disadvantages. Males and females have similar perspectives, except for two questions. For instance, females believe ChatGPT can be used for personal experience questions. The primary limitation of this study is that it was conducted in the initial phases of ChatGPT. It is anticipated that students will become more acquainted with this tool and other text-generation tools. Educators will explore ways to incorporate ChatGPT into students' required assignments. Future studies could compare AI tools and their impacts on educational fields. Subsequent studies could delve deeper into the potential of such tools for learning.

6 REFERENCES

[1] D. O. Eke, "ChatGPT and the rise of generative AI: Threat to academic integrity?" *Journal of Responsible Technology*, vol. 13, p. 100060, 2023. https://doi.org/10.1016/j.jrt.2023.100060

[2] S. S. Gill *et al.*, "Transformative effects of ChatGPT on modern education: Emerging era of AI Chatbots," *Internet of Things and Cyber-Physical Systems*, vol. 4, pp. 19–23, 2024. https://doi.org/10.1016/j.iotcps.2023.06.002

[3] A. Rejeb, K. Rejeb, A. Appolloni, H. Treiblmaier, and M. Iranmanesh, "Exploring the impact of ChatGPT on education: A web mining and machine learning approach," *The International Journal of Management Education*, vol. 22, no. 1, p. 100932, 2024. https://doi.org/10.1016/j.ijme.2024.100932

[4] A. Ammar, "The impact of ChatGPT on student's behavior in UAE," *International Journal of Theory of Organization and Practice (IJTOP)*, vol. 4, no. 1, pp. 1–9, 2024.

[5] C. K. Lo, "What is the impact of ChatGPT on education? A rapid review of the literature," *Education Sciences*, vol. 13, no. 4, p. 410, 2023. https://doi.org/10.3390/educsci13040410

[6] N. Chomsky, I. Roberts, and J. Watumull, "Noam Chomsky: The false promise of ChatGPT," *The New York Times*, 2023. Accessed: Dec. 1, 2023. [Online]. Available: www.nytimes.com/2023/03/08/opinion/noam-chomsky-chatgpt-ai.html

[7] C. Stokel-Walker, "ChatGPT listed as author on research papers: Many scientists disapprove," *Nature*, vol. 613, pp. 620–621, 2023. https://doi.org/10.1038/d41586-023-00107-z

[8] W. Yang, "Artificial Intelligence education for young children: Why, what, and how in curriculum design and implementation," *Computers and Education: Artificial Intelligence*, vol. 3, p. 100061, 2022. https://doi.org/10.1016/j.caeai.2022.100061

[9] M. Abdullah, A. Madain, and Y. Jararweh, "ChatGPT: Fundamentals, applications and social impacts," in *2022 Ninth International Conference on Social Networks Analysis, Management and Security (SNAMS)*, Milan, Italy, 2022, pp. 1–8, 2022. https://doi.org/10.1109/SNAMS58071.2022.10062688

[10] J. Rudolph, S. Tan, and S. Tan, "ChatGPT: Bullshit spewer or the end of traditional assessments in higher education?" *Journal of Applied Learning and Teaching*, vol. 6, no. 1, pp. 342–363, 2023. https://doi.org/10.37074/jalt.2023.6.1.9

[11] M. M. Rahman and Y. Watanobe, "ChatGPT for education and research: Opportunities, threats, and strategies," *Applied Sciences*, vol. 13, p. 5783, 2023. https://doi.org/10.3390/app13095783

[12] G. Cooper, "Examining science education in ChatGPT: An exploratory study of generative artificial intelligence," *Journal of Science Education and Technology*, vol. 32, pp. 444–452, 2023. https://doi.org/10.1007/s10956-023-10039-y

[13] E. S. Kasneci, "ChatGPT for good? On opportunities and challenges of large language models for education," *Learning and Individual Differences*, vol. 103, p. 102274, 2023. https://doi.org/10.1016/j.lindif.2023.102274

[14] S. Atlas, "ChatGPT for higher education and professional development: A guide to conversational AI," University of Rhode Island: Kingston, RI, USA, 2023. Accessed: Dec. 1, 2023. [Online]. Available: https://digitalcommons.uri.edu/cba_facpubs/548

[15] S. S. Gill and R. Kaur, "ChatGPT: Vision and challenges," *Internet of Things and Cyber–Physical Systems*, vol. 3, pp. 262–271, 2023. https://doi.org/10.1016/j.iotcps.2023.05.004

[16] M. Hutson, "Could AI help you to write your next paper?" *Nature*, vol. 611, pp. 192–193, 2022. https://doi.org/10.1038/d41586-022-03479-w

[17] Y. Wardat, M. A. Tashtoush, R. AlAli, and A. M. Jarrah, "ChatGPT: A revolutionary tool for teaching and learning mathematics," *Eurasia Journal of Mathematics, Science and Technology Education*, vol. 19, pp. 1–18, 2023. https://doi.org/10.29333/ejmste/13272

[18] A. Michalak, "Tools such as ChatGPT threaten transparent science; here are our ground rules for their use," *Nature*, vol. 613, no. 7945, p. 612, 2023. https://doi.org/10.1038/d41586-023-00191-1

[19] M. Sullivan, A. Kelly, and P. McLaughlan, "ChatGPT in higher education: Considerations for academic integrity and student learning," *Journal of Applied Learning & Teaching*, vol. 6, no. 1, pp. 31–40, 2023. https://doi.org/10.37074/jalt.2023.6.1.17

[20] M. Halaweh, "ChatGPT in education: Strategies for responsible implementation," *Contemporary Educational Technology*, vol. 15, no. 2, p. ep421, 2023. https://doi.org/10.30935/cedtech/13036

[21] V. Srinivasan, "AI & learning: A preferred future," *Computers and Education: Artificial Intelligence*, vol. 3, p. 100062, 2022. https://doi.org/10.1016/j.caeai.2022.100062

[22] J. K. Ali, M. A. Shamsan, T. A. Hazem, and A. A. Mohammed, "Impact of ChatGPT on learning motivation: Teachers and students' voices," *Journal of English Studies in Arabia Felix*, vol. 2, no. 1, pp. 41–49, 2023. https://doi.org/10.56540/jesaf.v2i1.51

[23] M. Firat, "What ChatGPT means for universities: Perceptions of scholars and students," *Journal of Applied Learning and Teaching*, vol. 6, no. 1, pp. 57–63, 2023. https://doi.org/10.37074/jalt.2023.6.1.22

[24] T. Al Shloul, T. Mazhar, M. Iqbal, Y. Yaseen Ghadi, F. Malik, and H. Hamam, "Role of activity-based learning and ChatGPT on students' performance in education," *Computers and Education: Artificial Intelligence*, vol. 6, p. 100219, 2024. https://doi.org/10.1016/j.caeai.2024.100219

[25] I. S. Chaudhry, S. A. Sarwary, G. A. El Refae, and H. Chabchoub, "Time to revisit existing student's performance evaluation approach in higher education sector in a new Era of ChatGPT—A case study," *Cogent Education*, vol. 10, no. 1, p. 2210461, 2023. https://doi.org/10.1080/2331186X.2023.2210461

[26] T. Rasul *et al.*, "The role of ChatGPT in higher education: Benefits, challenges, and future research directions," *Journal of Applied Learning and Teaching*, vol. 6, no. 1, pp. 1–16, 2023. https://doi.org/10.37074/jalt.2023.6.1.29

[27] A. Shoufan, "Exploring students' perceptions of ChatGPT: Thematic analysis and follow-up survey," *IEEE Access*, vol. 11, pp. 38805–38818, 2023. https://doi.org/10.1109/ACCESS.2023.3268224

[28] D. R. Cotton, P. A. Cotton, and J. R. Shipway, "Chatting and cheating: Ensuring academic integrity in the era of ChatGPT," *Innovations in Education and Teaching International*, vol. 61, no. 2, pp. 228–239, 2024. https://doi.org/10.1080/14703297.2023.2190148

7 AUTHORS

Mohammed Alnahhal is with the Mechanical and Industrial Engineering Department, American University of Ras Al Khaimah, Ras Al Khaimah, United Arab Emirates (E-mail: mohammed.alnahhal@aurak.ac.ae).

Hamda Alali is with the Department of Computer Science and Engineering, American University of Ras Al Khaimah, Ras Al Khaimah, United Arab Emirates.

Reema Alshamsi is with the Chemical and Petroleum Engineering Department, American University of Ras Al Khaimah, Ras Al Khaimah, United Arab Emirates.

Online-Journals.org

iJET International Journal of Emerging Technologies in Learning

iJET | eISSN: 1863-0383 | Vol. 19 No. 6 (2024) | OPEN ACCESS

https://doi.org/10.3991/ijet.v19i06.48591

PAPER

Designing, Coding and Embroidering: A Workflow for Gender-Sensitive and Interdisciplinary Teaching

Bernadette Spieler[1,2](✉), Sarina Gursch[3], Vesna Krnjic[2], Karin Horneck[4], Wolfgang Slany[2,3]

[1]Zurich University of Teacher Education, Zurich, Switzerland

[2]Catrobat Association, Graz, Austria

[3]Graz University of Technology, Graz, Austria

[4]bits4kids OG, Graz, Austria

bernadette.spieler@phzh.ch

ABSTRACT

Research suggests that girls' initial interest in computer science tends to decline during their teenage years, a trend that is not observed among boys. This paper addresses this gender gap and proposes integrating programming into handicraft lessons to provide a creative activity for all students. The project was conducted in three Austrian schools over one year (2019–2020), involving 229 middle school students. The evaluation included questionnaires and the assessment of programmed and stitched designs to structure gender-sensitive workshops. While boys consistently reported higher scores than girls in interest, sense of belonging, and enjoyment, girls' scores remained more stable. However, girls were significantly more likely to express pride in their final designs than boys, and overall, more girls completed individual designs as final products. These findings can be applied to interdisciplinary handicraft lessons in line with the Maker-Education movement to foster interest in programming.

KEYWORDS

visual programming, Embroidery Designer, computational handicraft, female teenagers

1 INTRODUCTION

Skills related to digitalization are becoming increasingly important worldwide [1]. New technologies such as the Internet of Things, robotics, wearables/augmented reality, artificial intelligence, blockchain, or big data are leading to new products, production processes, and distribution channels. Ultimately, these technological developments are resulting in new occupational fields, activities, skills, and qualifications as well. Technology and disruption will change every industry in the future. According to the WEF study [2], "The Future of Jobs 2018," positive job growth is expected in the coming years; simultaneously, the quality, location, format, and duration of new jobs will change significantly. Jobs will shift between continents (more jobs in Asia and the US), and depending on the technology acceptance and adaptability of the workforce, the impact will vary across industries. Women in

Spieler, B., Gursch, S., Krnjic, V., Horneck, K., Slany, W. (2024). Designing, Coding and Embroidering: A Workflow for Gender-Sensitive and Interdisciplinary Teaching. *International Journal of Emerging Technologies in Learning (iJET)*, 19(6), pp. 79–104. https://doi.org/10.3991/ijet.v19i06.48591

Article submitted 2024-02-16. Revision uploaded 2024-04-19. Final acceptance 2024-04-19.

Europe, in particular, will be strongly affected, as they are often employed in commercial occupations that will be under pressure due to digitalization. STEM subjects (Science, Technology, Engineering, and Mathematics) are professions that are gaining importance with digitalization. These are precisely the fields where women are underrepresented (the share of women working in STEM in Europe currently is 18.4%; [3], [4]). Worldwide, 36% of women work in the "technology, information, and media" sector [5]. In addition, many professions associated with commercial apprenticeships today are among those at risk from digitalization, data science, and automation (see WEF Gender Gap Report Chapter 3, "Gender Gaps in Jobs of Tomorrow" [2]).

While the number of women in some STEM disciplines, such as mathematics or the natural sciences, is slowly increasing, women continue to be severely underrepresented in the information and communication technology (ICT) sector. Women are underrepresented among information technology professionals in Central and Eastern Europe [6]. This contrasts with overall employment, where men and women are largely equally represented. Furthermore, a study in Switzerland [7] examined the career aspirations of boys and girls at age 15 and again at age 21, revealing the following trends: Young men at the age of 21 tend to aspire to higher status and more prestigious professions than those at the age of 15. About half of these career aspirations are STEM professions, such as computer scientist, electrical engineer, or architect. In contrast, most of the career aspirations of young women change only slightly, and among the ten occupations mentioned, there is practically no STEM occupation, such as primary school teacher, doctor, health care worker, or kindergarten teacher. This indicates that the advantages of occupational options are evaluated differently by boys and girls, leading to gender-typical occupational decisions. These gender differences in interests, self-concepts, and self-assessment of one's own abilities and skills become apparent early on, between 10 and 15 years of age [8].

But why are women essential in the field of ICT? According to a report by the European Commission [1], there are clear signs of unmet demand for skilled workers in most IT occupations. In particular, the large number of vacancies combined with low unemployment rates and the heavy reliance on foreign workers indicate that meeting the demand for skilled workers is challenging. High qualification requirements and the associated high specificity of IT occupations make it even more challenging to meet the demand for appropriately qualified skilled workers.

A possible solution could be to introduce children to these topics in an engaging way as early as possible. Competencies such as gathering information, problem-solving, and the use of tools to create digital content, such as programming, could be integrated into the classroom at early levels. Currently, this is only done to a limited extent in most countries. As a result, computer science (CS) remains largely unknown to most students worldwide. Creative subjects such as textile design or arts and crafts offer significant potential to change girls' attitudes toward STEM subjects, especially. Due to prevailing stereotypes, many girls do not feel comfortable, motivated, or capable of pursuing careers in these fields. Therefore, we need to broaden girls' horizons and make promising and creative STEM professions more accessible to them.

Our work builds on many years of gender-specific study on the impact of various programming activities using the Pocket Code App on girls' interest in learning programming [4], [8–13]. This project aims to develop a study-based, gender-sensitive curriculum for programming and stitching courses to encourage broader participation in programming activities. This is achieved by addressing structural inequalities

and beliefs that have historically limited the involvement of marginalized groups, such as women, in technology. Additionally, the project aims to foster a gender-sensitive mindset among teachers and students to ensure equal opportunities for all students. By providing female students with access to quality CS education in a predominantly male-dominated field, the project aims to enhance their engagement and skills.

In this paper, we aim to address the following research questions:

- Do stitching and coding activities promote intrinsic motivation, especially in girls (e.g., interest, self-efficacy, sense of belonging, fun, and engagement)?
- Are there differences in the design and pattern creation processes in stitching and coding courses between genders?

Furthermore, supporting materials were developed based on our findings from interviews, observations, and final artifacts, which specifically supported the second cycle (during the COVID-19 pandemic).

To address the research questions, our paper is structured as follows: Section 2 summarizes the findings from the literature on girls in CS, both in and out of school. Section 3 introduces the "Code'n'Stitch,"-project the method, and the associated study involving 229 middle school students. Section 4 presents the findings, which are further discussed in Section 5. Finally, Section 6 concludes the paper and offers an outlook.

2 RELATED WORK

In recent years, there have been numerous individual projects and initiatives explicitly targeting young girls. An interdisciplinary study is investigating various approaches to positively change the current situation of women in technology. In these studies and projects, several characteristics have been repeatedly highlighted and will be addressed in this section.

However, it is important to emphasize that the issue of women's underrepresentation in STEM subjects is not caused by a lack of interest or skills among girls and women [14]. Instead, it is systemic barriers that hinder them from reaching their full potential in these fields. For instance, stereotypes labeling STEM subjects as "masculine" can discourage girls and women, eroding their confidence in their capabilities in these fields [15].

2.1 Different (low-threshold) offers for different age groups (out-of-school activities)

Choosing a career path is a multifaceted decision, often influenced by various factors or a continuum of events. Although some of the factors in the decision-making process are obviously beyond the control of teachers, such as students' backgrounds (household income, ethnicity, social environment) or the views and preferences of family members, many other factors can be addressed through student counseling programs. Different studies have concluded that the most significant results came from programs that had been running for several years at the undergraduate level [16], [17]. Early exposure to CS can strengthen self-efficacy beliefs, academic performance, and foster interest [16], [18], [19]. This has already been demonstrated in the literature study by the authors of this paper [18].

In addition, the findings of [20] assessed the long-term impact of a series of extracurricular outreach events that introduced students to the discipline of computing, nurturing creative computational thinking through problem-solving and game programming. Overall, their study shows that extracurricular programs for girls have a positive impact on their future educational career decisions. Their confidence in studying CS was further strengthened, and their attitude toward CS changed positively. Interviews and general categories were summarized in the findings to describe the long-term effects of an extracurricular computer event. Based on 20 interviews (2–5 years after the intervention), four general categories were summarized to illustrate the long-term impacts of an extracurricular computer course. First, six participants felt that the intervention confirmed their decision to pursue a career in CS. They appreciated the opportunity to create something new and connect games to their own identities. Second, for five participants, the intervention led them to consider CS as a possible career option. The new information and experiences they received from the intervention, as well as their physical presence in a university building, were instrumental in their decision to study CS. Third, for six participants, the intervention had no significant impact on their career plans. They did not choose CS as a major, although they participated in the intervention for a variety of reasons, such as a desire to experiment or the benefit of being able to create their own website. Finally, for three participants, the intervention confirmed that CS was not an interesting career option for them. However, the course helped them understand programming and dispel some of their misconceptions about this field.

2.2 Gender-sensitive computer science education

Extracurricular initiatives are not enough. Children may be excluded for various reasons (e.g., children in rural areas, high costs, or a one-sided focus on a specific target group, e.g., those with prior knowledge). While in some countries, CS is already a compulsory subject starting in Grade 5 (e.g., in different regions in Germany [21] and Switzerland [22]), in other countries like Austria, it is mostly an activity integrated into subjects [23]. Therefore, it is also necessary to encourage teachers to introduce various CS content into the classroom in a playful way and at an early age. This paper [19] summarizes the results of more than 800 publications to provide educators with a comprehensive and easy-to-navigate map of interventions. They came up with the following umbrella strategies:

- Use inquiry-based and real-world learning activities to engage students in computing.
- Showcasing as many facets and interdisciplinary applications of CS as possible, as early as possible, to attract students from diverse disciplines.
- Split classes based on experience, gender, or shared interests for optimal results.
- Give more emphasis to the process of thinking, designing, and problem-solving than to the actual programming.
- Use visual programming environments to teach introductory programming.
- Take students to events and excursions and share stories and role models from the history of computing with them.

In addition, the articles examined summarize efforts to increase girls' interest in CS. Six key aspects were identified to make CS more engaging for this target group:

combating false stereotypes, motivating and sparking initial interest, providing appropriate first exposure, creating a less hostile learning environment, fostering self-confidence, and sustaining long-term interest.

2.3 Highlighting the creative and multi-faceted nature of computer science activities

The study of [24] is an essential step toward understanding how to spark girls' interest in STEM. They implemented an after-school program for middle school girls that integrated narrative-based, blended learning, and design-based engineering activities. This initiative, which used design challenges and mentors from diverse backgrounds, appeared to increase girls' awareness of STEM skills, identities, and interests. A study by [25], building on the importance of fostering interest in STEM, also highlighted that a broad understanding and personal interest in CS are critical to girls' participation. In this context, teachers play a central role in fostering girls' CS aspirations, indicating the importance of the educational environment. When considering the characteristics of problem-solving strategies, [19] found that girls and boys approach problem-solving differently. While girls tend to define problems more broadly, boys tend to approach problems in isolation. This broad approach can put more pressure on girls as they strive to consider every detail. Other authors suggested that an effective strategy might be a mix of both approaches [26]. In addition, the holistic thinking attributed to girls and women is related to the concept of computational thinking (CT), an important aspect of our technology-driven society. CT is about organizing extensive information into complex patterns and identifying connections—competencies attributed to girls and women. However, it is critical to note the apparent underrepresentation of a multidisciplinary approach in CS education. This is related to the issue of student engagement, where the lack of challenge, skill, and relevance in the classroom can lead to disinterest. This idea is consistent with Csikszentmihalyi's concept of flow, which states that task difficulty should be carefully balanced to promote engagement [27]. In essence, these studies shed light on the various facets of promoting greater engagement of girls in STEM subjects and the potential strategies that should be explored.

Buhnova and Happe summarized a literature review on practices to create girl-friendly CS teaching [28]. Additionally, they gathered insights from their own experiences teaching girls in their courses. They identified the following key points:

1. Creating safe environments: It is crucial for girls to feel a sense of belonging. This can be achieved when they feel understood and when the goals and activities are meaningful and relevant to them.
2. Segregation: to provide all students with a fair share of instruction time and suitable instructional form.
3. Working in teams: Collaboration and teamwork can help increase girls' engagement and participation in CS classes (if organized appropriately).
4. Personalized learning: for example, by offering self-assessment interventions through encouragement and feedback to reduce frustration.

The fact that girls prefer a safer environment without pressure and competition and with more time for their tasks to be finished could be misinterpreted by the teacher as these girls being weaker, which is not necessarily the case [29].

Given the existing study on gender-sensitive pedagogy, there is still a need to understand the specific effects on girls' intrinsic motivation in STEM subjects, such as interest, self-efficacy, sense of belonging, and engagement (e.g., through playful elements). Our first proposed research question aims to address this gap and provide a more nuanced examination of how specific activities can foster positive attitudes and experiences for girls in science, technology, engineering, and mathematics.

2.4 Prior work with code and stitch courses

The concept of programming patterns is not new. The "TurtleStitch" project (https://www.turtlestitch.org/; [29]), based in Austria/Vienna, introduced this idea at the Scratch conference in 2015. Moreover, the Maker-Education movement has also recognized these possibilities [11]. Through the advancement of various physical computing circuits, computing fashion can be made wearable and interactive [30]. For instance, creating embroidered fabrics can be enhanced by incorporating conductive threads or LED lights, transforming them into e-textiles or even "smart wearables."

To sum up, programming patterns are widely used to teach programming to young people through textiles and wearables [31], [32]. Some researchers also concentrate on strategies to engage girls in technology by utilizing smart textiles or embroidery programming [4], [33].

Existing research lacks a detailed exploration of gender differences in design and pattern creation processes within stitching and coding courses. Understanding whether end products differ based on gender, either through adaptation during the project or from tutorial-based developments, is uncharted territory. The second proposed research question aims to address this gap and provide deeper insights into gender dynamics within these courses.

3 MATERIALS AND METHODS

3.1 The "Code'n'Stitch"-project

The "Code'n'Stitch" project was a two-year investigation (Sep. 2018–Sep. 2020) with the aim of testing digital pattern-making as an interdisciplinary method for learning programming in middle schools (Grade 5–7; https://catrob.at/codeNstitch). The project was a collaboration between Graz University of Technology (TU Graz; responsible for the app development), bits4kids (https://www.bits4kids.at/; in charge of the workshops and creating the learning cards and tutorials), the Styrian University of Teacher Education (providing didactic expertise), and the fashion shop "Apflputzn" (https://www.apflbutzn.at/), which introduced the topic of bio-fair fashion to schools and supplied bio-fair shirts and bags for stitching. This project was funded by FEMtech (https://femtech.at), an organizational unit of the Austrian Federal Ministry for Climate Protection, Environment, Energy, Mobility, Innovation, and Technology (BMK), which aims to empower women in study and technology and promote equal opportunities. Therefore, a special focus was placed on the gender-sensitive design of the workshop. For instance, at the project's outset, a gender equality and diversity course was conducted with all project members and trainers who would lead the courses in schools. The material was prepared attractively and aesthetically, dialogues on role models were initiated (e.g., through course cards; see Figure 1),

emphasis was placed on problem-solving (e.g., abstracting desired patterns), great importance was given to the design and creative process (two lessons were dedicated to designing), and a unique visual programming language was developed. The app and materials were developed by observing our target group, creating personas based on this, and conducting initial workshops for a needs analysis of the app in 2018/19 [11]. Thus, many of the points mentioned in Section 2 were taken into account.

Fig. 1. An example of a course card for encouraging dialogues and fostering role models (First cycle, 2019)

The focus of the project was to: 1) provide guidance and support to teachers without programming skills; and 2) show girls, in particular, a new way to express themselves creatively through programming. Furthermore, the Embroidery Designer App ([12]; https://catrob.at/ED) was developed. This is a new version of the Pocket Code App from the Catrobat project ([34]; https://catrobat.org) at TU Graz. The Pocket Code App offers a visual programming language where users can develop games, animations, interactive music videos, control drones, and more using smartphones or tablets. Similar to Scratch (https://scratch.mit.edu/), the Pocket Code App utilizes bricks as the foundation for creating program code. New bricks were created to expand their functionality specifically for this purpose. For instance, the needle brick allowed users to stitch and control embroidery machines. This version offers the possibility to create embroidery files with special blocks that conventional embroidery machines can execute. Besides these new possibilities that programming offers for handicraft lessons, it should be noted that handicraft lessons are still very traditionally oriented in many schools, and digital tools such as laser cutters are only slowly being introduced [35]. Especially within textile lessons, there are still very few possibilities, and the teachers usually do not have the necessary knowledge to combine this subject with CS content. The project is, therefore, especially aimed at teachers without prior programming knowledge.

3.2 Context and participants

We conducted our activities in the handicraft classes of three different schools in Graz, Austria. The workshops were divided into two cycles: September 2019 to

January 2020 (Cycle 1) and March to June 2020 (Cycle 2). Two classes finished the cycle 2 workshop in March 2020, before the onset of the Corona pandemic and subsequent school closures. Due to the pandemic, adjustments had to be made from March 2020 onwards. This paper primarily focuses on the first cycle, as some evaluations, such as questionaries, were discontinued due to the shift to online or hybrid classes. Nonetheless, many new insights were also obtained through online courses, which are detailed at the end of the paper. Figure 2 depicts the contexts of both cycles.

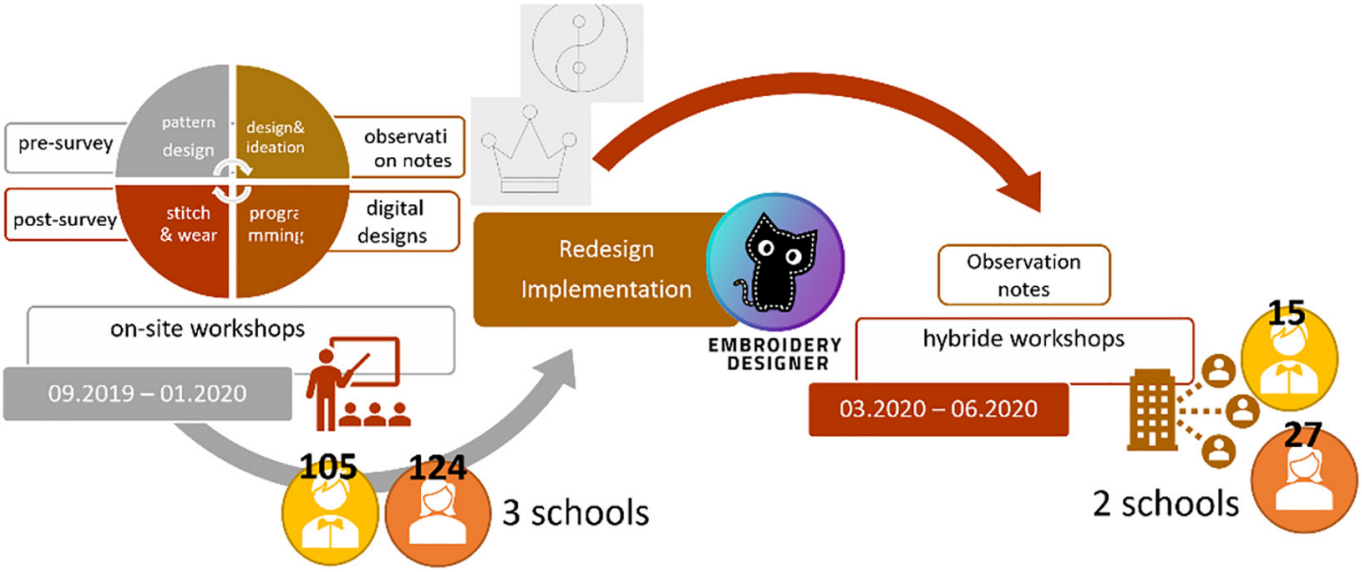

Fig. 2. Study design for cycle 1 (pre-Corona) and cycle 2 (post-Corona)

3.3 Workshops

The workshops during the first cycle were based on eight 50-minute lessons, summarized in four units. In these units, students went through four open-ended and interest-driven phases: 1) the pattern design phase; 2) the design and ideation phase; 3) the programming phase; and 4) the stitch and wear phase (see Figure 3). In the last unit, the whole team was present and helped students embroider their designs with programmable embroidery machines in schools. Additionally, some students finished their designs independently.

In the following, the different phases are described in more detail: During the pattern design phase, the students discussed with the teacher different types of stitches, fabrics, and other information related to embroidery that they will come across when working on textile projects. Then the student groups created their own designs based on initial internet research or their own ideas. In this step, students were motivated to think about what geometric shapes are used in their design, what angles are needed, as well as other mathematical features of their design (e.g., formulas). This phase was intentionally placed before the programming unit to not limit students' ideas and designs from the beginning. Students received approval and tips from the course instructors (trainers from bits4kids or the TU Graz team) on whether their sketches constituted a "programmable" design. For example, designs that contain too many details, curves, or completely flat fills are considered complex.

In contrast, good embroiderable designs are lines, circles, or other geometric figures. For this, students started to draw their desired design on a piece of quadratic paper. During the programming phase, students received instructions on important control structures in programming (e.g., loops, conditions, messages, etc.). After this, students had four lessons on programming with the Pocket Code App before they stitched their designs on fabrics (stitch and wear phase). In these units, failure and overcoming unforeseen challenges are expected and part of the learning process [36].

Fig. 3. The design – code – stitch workflow

229 students participated in the workshops during the first cycle. Table 1 shows the gender distribution across the workshops in relation to the individual schools.

Table 1. Distribution and grades/gender of participating schools/students during the first cycle

	Female	Male	Total
	123	106	229
School 1 (secondary school, Grade 7, students between 12 to 13 years old)	34	44	78
School 2 (grammar school, Grade 6, students between 11 to 12 years old)	56	28	84
School 3 (middle school, Grade 5–7, students between 10 to 14 years old)	33	34	67

In these workshops, we use a prototype version of the app. At that time, there was only one command available to create patterns: a single stitch. Although all possible patterns could already be created using this single command, we recognized the necessity of incorporating more nested commands for common use cases.

The second cycle of the project was initially planned to work in the same three schools with the same classes and students. However, due to school closures resulting from the COVID-19 pandemic, the study design had to be significantly revised. Nevertheless, in the conclusion of the paper, we briefly outline the main results that have emerged during these modifications. Building on the findings from the first cycle, additional features were incorporated into the app (e.g., more stitches), along with tutorials inspired by students' suggestions (refer to Figure 4).

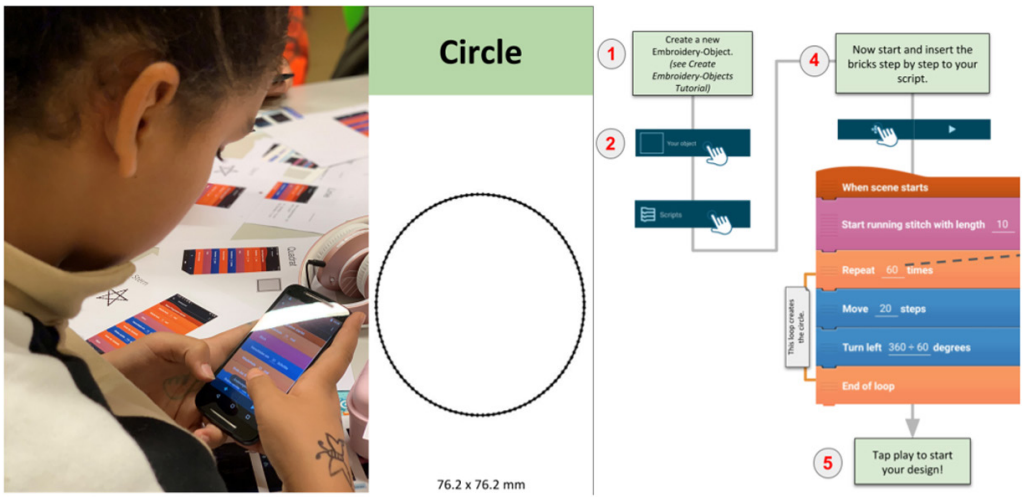

Fig. 4. Tutorial for creating a circle, step-by-step

At the end of the second cycle, a wide range of tutorials for various patterns (easy to medium) were made available, such as stars, flowers, or a Minecraft creeper figure. The second cycle aimed to validate the developed tutorials and the adaptation of the app in the school context based on the students' experiences. After the second cycle, the following resources were created: Following, a Wiki page (see Figure 5) will serve as a knowledge repository for all the materials we created. Second, an Instagram account where we posted new tutorials, photos, and videos online daily. Third, we created YouTube videos with short beginner tutorials and linked them to the wiki pages.

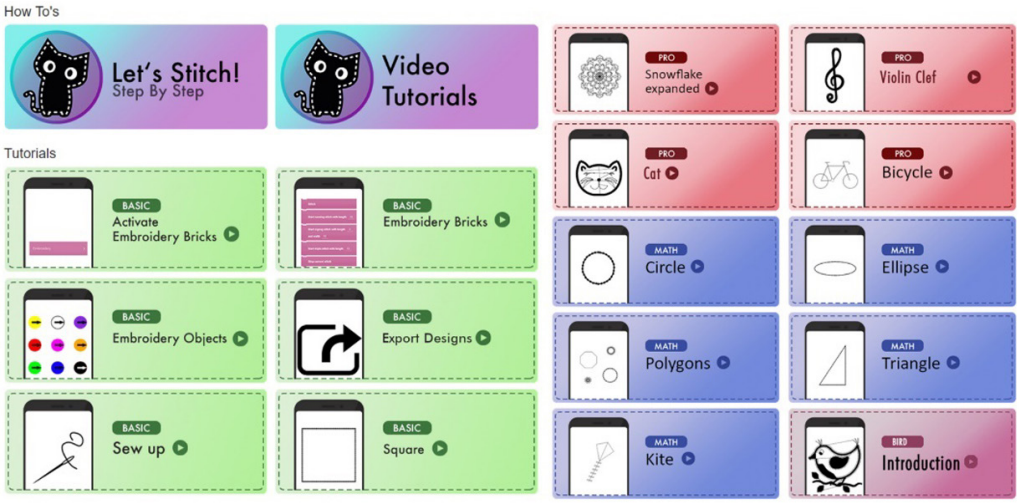

Fig. 5. Tutorial on Catrobats Wiki page

3.4 Data collection and analysis

Following permission from the school and teacher to conduct the research, the legal guardians of the children were informed of the research and signed their consent. For anonymization, we used individual codes for the students that were written in the surveys and used for the name of the programming patterns to be uploaded. These alterations have not distorted the scholarly meaning. The anonymized data made available to the public cannot be used to identify individuals. The data that support the findings of this study are openly available [37], [38]. Under these

circumstances, data collection and analysis proved to be rather difficult. During the first cycle, a total of 229 students (123 girls and 106 boys) aged between 10 and 14 participated in the questionnaire. Some problems were also encountered during the first cycle. For example, the pre- and post-project questionnaires were filled out differently per class. Some students filled out both questionnaires as planned (n = 110), others only the pre-questionnaire (n = 51), and still others only the post-questionnaire (n = 55; 14 pupils accordingly did not fill out either questionnaire). The questionnaire was used to ask about intrinsic motivators considered important for girls in CS, such as interest, self-efficacy, sense of belonging, and fun [8]. A 4-point Likert scale [39] was used to measure the variables that refer to 1) strongly disagree, 2) disagree, 3) agree, and 4) strongly agree. The questions have been developed on the basis of literature [40], the CATS Attitude Scale Items [41], and other studies [42]. No questions have been asked that could foster stereotype threats, as proposed by [41], e.g., "Girls can do technology as well as boys." To demand their attention, using no neutral value and counter questions is recommended [43]. Thus, it is not always "the higher, the better." Furthermore, prior knowledge of programming was only inquired about in the pre-questionnaire, while their experiences during the workshops were only addressed in the post-questionnaire [44]. Therefore, our questionnaire is being validated for the first time in this paper and tested for the reliability of all items on the scale. IBM SPSS Statistics was utilized for data analysis. Initially, frequencies for the four intrinsic parameters were calculated to gain more insights into the distribution and frequency of each value in the corresponding variables. This process aids in developing an understanding of the data and identifying potential problems or outliers. Subsequently, a reliability analysis is conducted to assess the internal consistency of the scales in the questionnaire. This step involves calculating statistics such as Cronbach's alpha to determine if the individual items on a scale are coherent with each other. A t-test is then employed to determine gender differences and examine possible variations between genders in relation to the means. Finally, an analysis of covariance, also known as an ANOVA (analysis of variance), is used to analyze potential differences between groups over time. This is achieved by calculating the means to generate summary statistics for each scale or intrinsic parameter. To explore the practices students used in creating their designs, as well as the challenges, situations, questions, and uncertainties that arose, data for analysis was gathered from field notes in the classroom [45]. Subsequently, the programmed and stitched designs were collected. For the analysis, design sketches were compared with the final products (programs) to determine if students were able to realize their original designs, if they created a tutorial instead, or if they had to make other adjustments. These programs were analyzed by gender and by patterns. A total of 217 final products (DST files) were collected from the first cycle.

This multiple-data-source approach provides a more comprehensive picture of classroom activities and student participation. However, the evaluation of the second cycle is incomplete due to the inability to distribute post-questionnaires and the unavailability of all program files. Consequently, only the experiences from the remote units are detailed for analysis to be utilized in future online and hybrid formats.

4 RESULTS

4.1 Evaluation of the questionnaires

For the evaluation of the intrinsic parameters of the questionnaire, the following datasets were utilized [37]. The initial question inquired about students'

programming experiences. Out of 112 students, 76.47% of girls and 64% of boys reported having no programming experience. All analyses were conducted at a significance level of less than 5% ($p < 0.05$).

Item interest. The first block of questions on interest included the following: "1a: Learning about programming interests me."; "1b: I find it important to have an idea about programming."; "1c: I find computer science and programming boring."; and "1d: I like programming."

The items in the questionnaire designed to assess "interest" demonstrate reliable measurement properties. The questions utilize a unidimensional scale, and the distribution of individual items follows a normal distribution, indicating good data quality. The Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy is significant (T1: 763, T2: 797), suggesting that the data is suitable for factor analysis.

Moreover, the overall explained variance exceeds 1 (T1: 2.546, T2: 2.802), indicating that the derived factors capture a substantial portion of the information in the data, supporting construct validity. Additionally, the reliability assessment yielded a Cronbach's alpha coefficient of T1: .807 and T2: .857 for the four items, indicating acceptable to good internal consistency.

Overall, the pre-questionnaire mean score on interest in programming indicates a general agreement towards having an interest (m = 3.20, sd = .643). After the intervention, the post-questionnaire mean decreased slightly (m = 3.04, sd = .729), showing a small decline in interest level with increased variation in responses. When the data are examined separately by gender, the following picture emerges: The results show that the interest between T1 (m = 3.19, sd = .63) and T2 (m = 3.07, sd = .70) tended to decrease slightly over time (t (109) = 1.69, p = .094). To test whether gender moderates interest in programming, a repeated-measures analysis of variance was conducted. Figure 6 shows that interest has declined only among boys. Although the interaction effect is not significant (F (1, 107) = 3.786, p = .054), boys' interest at T1 (m = 3.35, sd = .73) compared to that of girls (m = 3.08, sd = .55) was significantly higher (t (107) = 2.19, p = .031).

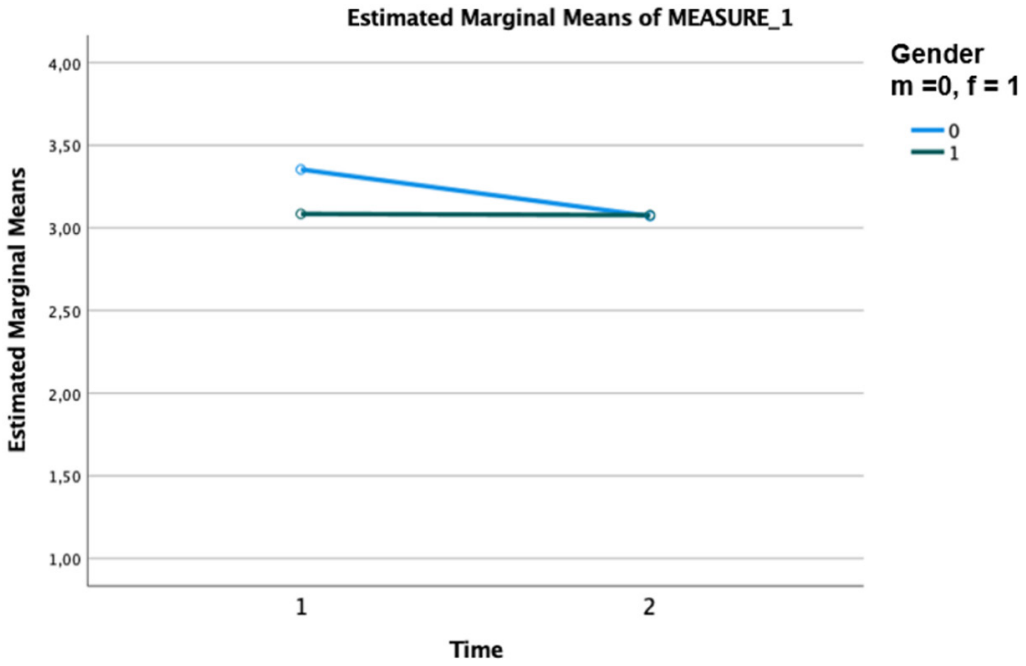

Fig. 6. Change in interest in programming, differentiated by gender

Item self-efficacy. The second block of questions on self-efficacy included the following: "2a: I generally feel confident in using my smartphone or PC."; "2b: I can learn programming."; "2c: I have little confidence when it comes to informatics (computers, smartphones, programming)"; "2d: I try to find several ways to solve problems I encounter in everyday life"; and "2e: Programming reflects my way of thinking." The results of the items concerning "self-efficacy" present some statistical challenges. Firstly, the KMO measure shows a value in the acceptable range (above 5: T1: .614, T2: .676), indicating that the data are reasonably suitable for factor analysis. However, the value is slightly below the desired value (above .7), indicating a potential for random influences on the results and, therefore, could be improved.

Overall, it is a one-dimensional scale with a total variance explained greater than 1 (T1: 1.738, T2: 1.901). Secondly, the reported Cronbach's alpha values are below the generally accepted threshold of 0.7 (T1: .506, T2: .563), suggesting that the test items may not correlate well with each other when measuring the same construct or that the items may be interpreted differently by respondents.

The mean scores at T1 ($m = 3.05$, $sd = .472$) and at T2 ($m = 2.91$, $sd = .525$) suggest that the intervention may not have improved participants' sense of self-efficacy concerning programming and that the effect of the intervention may vary greatly across participants. The self-efficacy scores for the t-test show the following results: At T1, boys' mean score was slightly higher ($m = 3.107$, $sd = .497$) than girls' ($m = 3.007$, $sd = .450$), but this difference was not statistically significant ($p = 0.182$). At T2, both mean scores decreased, with girls ($m = 2.846$, $sd = .544$) decreasing more than boys ($m = 3.013$, $sd = .475$), but this difference was also not statistically significant ($p = .48$). However, a significant change in self-efficacy over time can be seen ($p = .004$), indicating an overall decrease regardless of gender, but no significant interaction was found between time and gender ($F (1, 107) = 1.005$, $p = .054$), indicating that the change in self-efficacy from T1 to T2 was not significantly different between boys and girls.

Item sense-of-belonging. The third block of questions on sense of belonging included the following: "3a: I can imagine what people in technical professions do."; "3b: I can imagine how programming works."; "3c: I can imagine that programming is important for my future profession."; "3d: A technical profession would suit me."; and "3e: I think I have the same or similar characteristics as people who program."

In the scale test of the "sense of belonging" item, the KMO measure shows that the data are well-suited for factor analysis (T1: .773, T2: .731). The scale is identified as unidimensional (total variance explained: T1: 2.601, T2: 2.667). In the reliability test of the item, Cronbach's alpha (T1: .760, T2: .771) indicates a good level of internal consistency for this item.

The mean scores for this item at T1 ($m = 2.47$, $sd = .659$) and at T2 ($m = 2.4$, $sd = .671$) suggest that while there is some variability in responses, it is not excessive. Responses tend to cluster reasonably around the mean, suggesting that most participants feel a moderate sense of belonging to the program. The results of the t-test examining gender differences on the item sense of belonging in programming show some notable differences: boys at T1 ($m = 2.655$, $sd = .693$), girls ($m = 2.328$, $sd = .572$). This indicates that, on average, boys reported a stronger sense of belonging in programming than girls (significance level $p = .001$).

Mean scores at T2 showed the following for boys ($m = 2.637$, $sd = .735$) and girls ($m = 2.268$, $sd = .585$). Once again, boys reported a significantly stronger sense of belonging to programming than girls ($p < .001$). Additionally, this time, the rate of change over time did not show a significant difference between the genders ($F (1, 107) = 1.940$, $p = .167$).

Item fun and games. The fourth and final related block included questions about fun and games: "4a: Programming is fun for me."; "4b: I like playing

computer games."; "4c: I like using game apps."; and "4d: When I play or program on the computer, I sometimes forget everything around me."

The statistics presented for this item indicate acceptable consistency and suitability for factor analysis. The KMO measure (T1: .675. T2: .677) indicates moderate suitability of the data, and the total scores are above 1 at both time points, indicating general compatibility. Cronbach's alpha values (T1: .709, T2: .683) are relatively good, although the value for T2 is lower than desired.

The mean values at T1 (m = 2.89, sd = .747) and at T2 (m = 2.91, sd = .728) show a stable central tendency and relatively low dispersion. The results of the t-test showed significant gender differences in this domain at both T1 and T2. At T1, boys reported a higher mean score (m = 3.218, sd = .632) than girls (m = 2.625, sd = .709), a difference that was statistically significant ($p < .001$). This difference increased very slightly at T2, for boys (m = 3.258, sd = .608) and girls (m = 2.670, sd = .702), also showing significance ($p < .001$). However, the interaction effect is not significant over time ($F(1, 106) = .973$, $p = .326$).

In summary, scores in all areas either remained stable or declined slightly after the intervention. Boys consistently reported higher scores than girls in the areas of interest, sense of belonging, and fun/games, with statistically significant differences in the categories of interest (at T1), sense of belonging, and fun/games. There were no significant gender differences in the self-efficacy domain. The scale results suggest that the questions for this item may need to be adjusted.

Further, an investigation was conducted to identify correlations with the interest in programming at both T1 and T2. Several moderate-to-strong correlations were found (see Figure 7). According to [46], the effect size is considered low if the value of r fluctuates around 0.1, medium if r fluctuates around 0.3, and large if r fluctuates more than 0.5.

Correlations

		Interest in Programming Pre	Interest in Programming Post	Self-efficacy in Programming Pre	Self-efficacy in Programming Post	Sense-of Belonging in Programming Pre	Sense-of Belonging in Programming Post	Fun and Games Pre	Fun and Games Post
Interest in Programming Pre	Pearson Correlation	1	.409**	.634**	.273**	.548**	.278**	.374**	.320**
	Sig. (2-tailed)		<.001	<.001	.004	<.001	.003	<.001	<.001
	N	161	110	161	110	161	110	161	109
Interest in Programming Post	Pearson Correlation	.409**	1	.363**	.534**	.461**	.602**	.171	.382**
	Sig. (2-tailed)	<.001		<.001	<.001	<.001	<.001	.075	<.001
	N	110	165	110	165	110	165	110	164
Self-efficacy in Programming Pre	Pearson Correlation	.634**	.363**	1	.482**	.497**	.390**	.310**	.258**
	Sig. (2-tailed)	<.001	<.001		<.001	<.001	<.001	<.001	.007
	N	161	110	161	110	161	110	161	109
Self-efficacy in Programming Post	Pearson Correlation	.273**	.534**	.482**	1	.411**	.611**	.271**	.443**
	Sig. (2-tailed)	.004	<.001	<.001		<.001	<.001	.004	<.001
	N	110	165	110	165	110	165	110	164
Sense-of Belonging in Programming Pre	Pearson Correlation	.548**	.461**	.497**	.411**	1	.630**	.431**	.326**
	Sig. (2-tailed)	<.001	<.001	<.001	<.001		<.001	<.001	<.001
	N	161	110	161	110	161	110	161	109
Sense-of Belonging in Programming Post	Pearson Correlation	.278**	.602**	.390**	.611**	.630**	1	.255**	.430**
	Sig. (2-tailed)	.003	<.001	<.001	<.001	<.001		.007	<.001
	N	110	165	110	165	110	165	110	164
Fun and Games Pre	Pearson Correlation	.374**	.171	.310**	.271**	.431**	.255**	1	.646**
	Sig. (2-tailed)	<.001	.075	<.001	.004	<.001	.007		<.001
	N	161	110	161	110	161	110	161	109
Fun and Games Post	Pearson Correlation	.320**	.382**	.258**	.443**	.326**	.430**	.646**	1
	Sig. (2-tailed)	<.001	<.001	.007	<.001	<.001	<.001	<.001	
	N	109	164	109	164	109	164	109	164

**. Correlation is significant at the 0.01 level (2-tailed).

Fig. 7. Correlations between T1 and T2

Given that four items—interest, self-efficacy, sense of belonging to programming, and fun/games—show many strong positive correlations at both T1 and T2, this suggests that these variables are interrelated and influence each other over the duration of the programming course. For example, the positive correlation between interest and self-efficacy at T1 ($r(161) = .64$, $p < .001$) and at T2 ($r(110) = .273$, $p = .004$) may indicate that children who feel more capable (higher self-efficacy) tend to show more interest in programming and vice versa. Therefore, encouraging and improving students' self-efficacy could promote their interest in programming. A strong positive correlation between sense of belonging in programming and the other variables suggests that as children's sense of belonging in programming increases, so does their interest at T1 ($r(161) = .548$, $p < .001$) and T2 ($r(110) = .461$, $p < .001$), self-efficacy at T1 ($r(161) = .482$, $p < .001$) and at T2 ($r(110) = .411$, $p < .001$), and the item fun and games at T1 ($r(161) = .630$, $p < .001$) and at T2 ($r(110) = .431$, $p < .001$). This may suggest that creating an inclusive, welcoming learning environment where students feel a sense of belonging may have a positive impact on their engagement and performance in programming. A strong correlation between fun/games and the other variables (e.g., interest in programming at T1 ($r(161) = .374$, $p < .001$)) suggests that incorporating playful elements into the programming course can potentially increase the other factors (except interest in T2 $r(110) = .171$, $p = .075$).

The post-questionnaire (T2) contained two additional questions ($n = 165$): "5a: I felt comfortable and taken seriously in the "Code'n'Stitch" project." and "I am proud of the design I created." Both questions aimed to gather more information about the classroom climate and participants' designs.

To determine significance, a t-test was utilized. Girls showed higher agreement with question 5a compared to boys, but this difference was deemed insignificant (girls $m = 2.8632$, boys $m = 2.71$, $p = .376$). On the other hand, the results for the second question were considered significant (girls $m = 3.588$, boys $m = 3.177$, $p = .003$). This indicates that girls significantly expressed more pride in their designs than boys.

4.2 Results of the sketches and final programs

At the end of cycle 1, 217 final products (DST files) were uploaded on the Catrobat sharing page, and all 217 programs were embroidered on t-shirts or bags. For this, various designs were created. Table 2 shows whether students created their own designs or used prepared tutorials [38].

Table 2. Tutorials or own designs (final product) divided by gender

		Female	Male
Category 1	Tutorials	36 of 116 (31%)	40 of 97 (41%)
Category 2	Own designs	80 of 116 (69%)	57 of 97 (59%)

Figure 8 shows the various shapes that have been created. These are screenshots from their final DST files. This standard stitch-based file format can be read and converted by many commercially available embroidery machines (e.g., machines by Brother or Bernina).

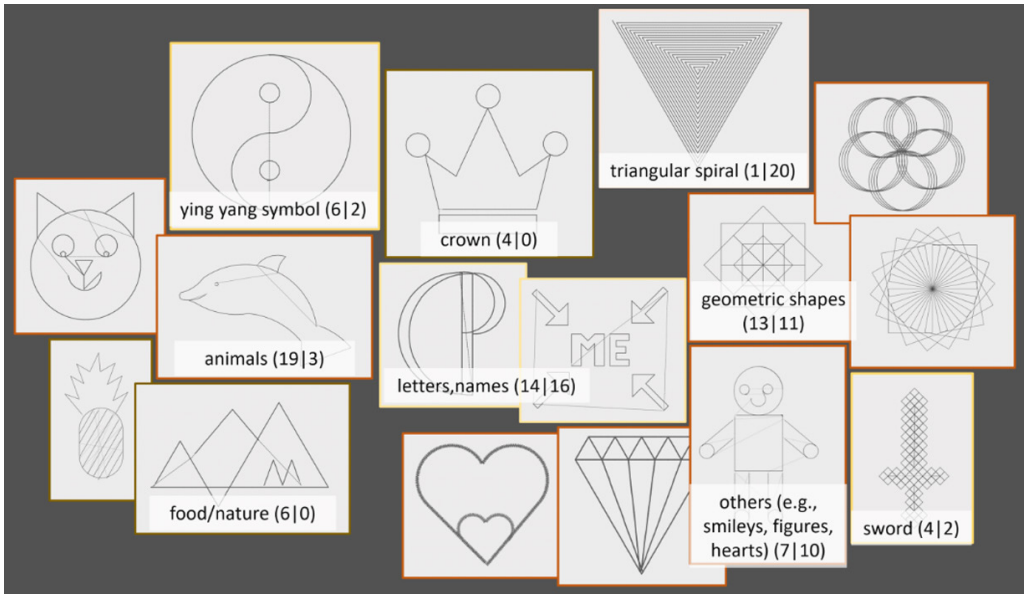

Fig. 8. Different patterns (final products) divided by gender (first number: female, second number: male students)

The initial sketches were used on-site during the units. Some were copied or collected, but not all were secured. A total of 81 (f = 47, m = 34) initial sketches that could be assigned to final products were collected. 35 students (f = 23, m = 12) implemented their final product based on their initial sketch, and 39 (f = 21, m = 18) had to change their idea during the programming. 7 (f = 3, m = 4) students could implement their initial sketch with modifications. Figure 9 shows examples of an initial template compared to the final product. Figures 10 and 11 display some examples of an idea sketch compared to the final product, which was changed during the workshop. Figure 12 illustrates an example of an initial template compared to the final product (the DST files) with modifications.

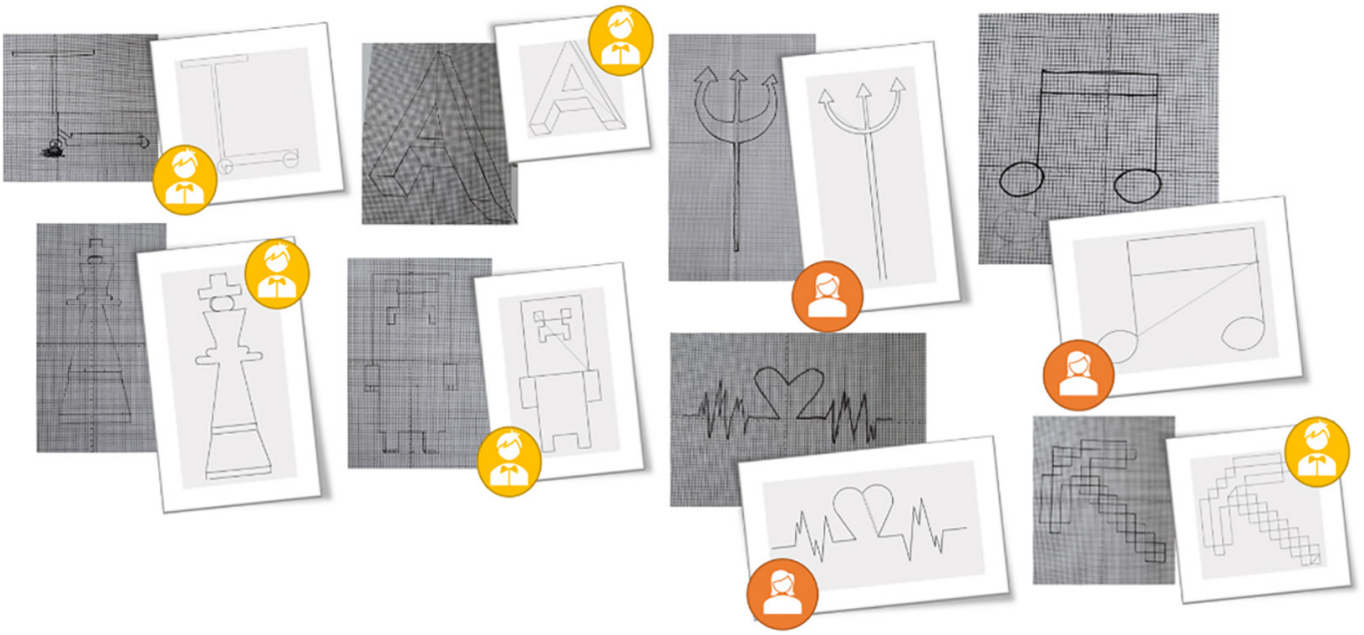

Fig. 9. Some examples of an initial template and final product (DST file)

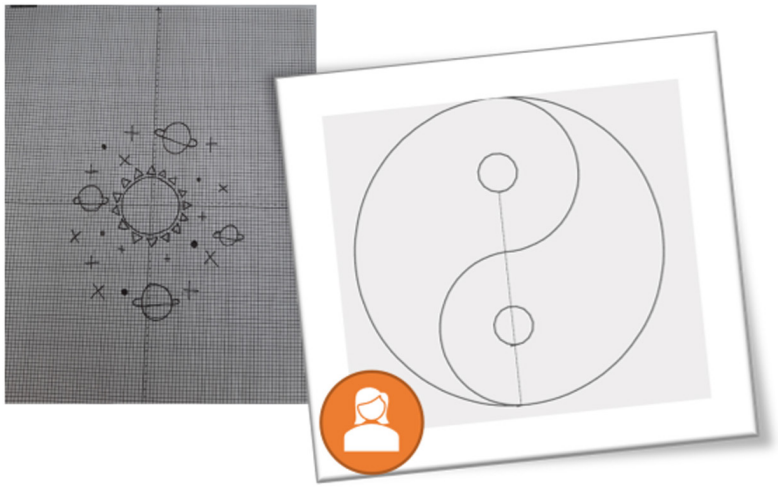

Fig. 10. Example of an initial template and changed the final product

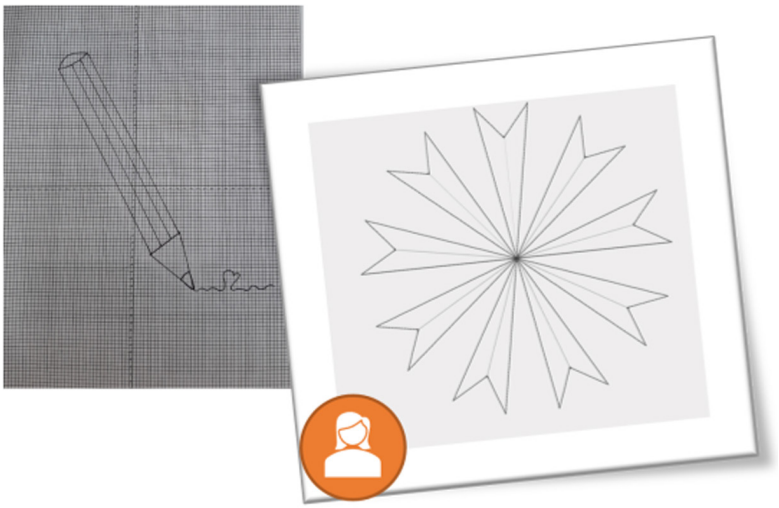

Fig. 11. Example of an initial template and changed the final product

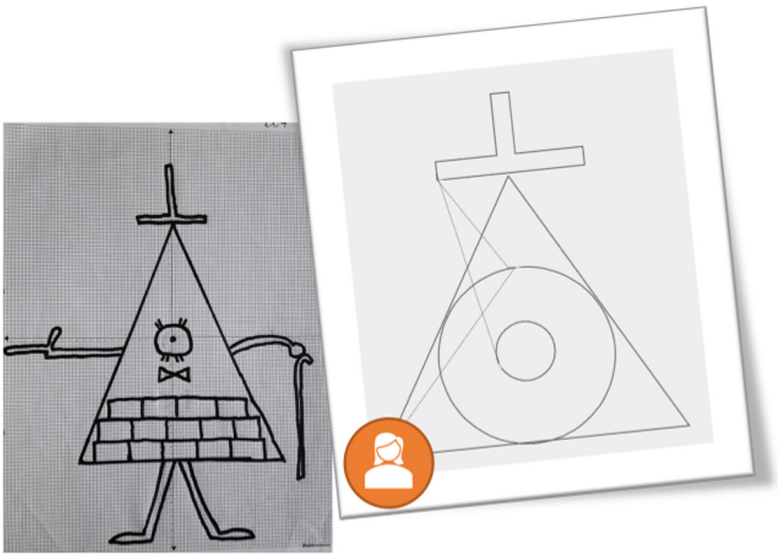

Fig. 12. Example of modifications of initial template to the final product

4.3 Evaluation of the field notes

The on-site observations during the first cycle revealed the following insights: The students encountered challenges in programming various patterns using only one stitch. They needed to quickly incorporate new stitches, such as the "ZigZag" stitch for wider lines or automatic stitches while moving the objects. Many of the patterns were too large for embroidery, requiring the use of external applications such as "Stitch Pro" to determine the final design size. It was observed that the lessons were more effective when the teachers actively participated in the teaching process. Students desired to showcase their designs to the teacher, emphasizing the importance of receiving feedback. When teachers were not engaged, children's motivation and enjoyment of the lessons decreased. In terms of gender-specific observations, boys displayed more confidence in programming but were quicker to seek help when faced with challenges. On the other hand, girls tended to start with tutorials and hesitated to create their own designs initially. However, they demonstrated greater persistence, were more likely to persevere, and produced more unique designs consistently. These observations were also reflected in the questionnaire.

5 DISCUSSION

The questionnaire conducted during the first cycle also reflected many of the problems with the app prototype. A limitation of the study is that the questionnaire was no longer conducted after the app was further developed to be used in cycle 2. Nevertheless, the pre- and post-evaluation of the questionnaire tells us a lot about gender-specific perceptions of intrinsic motivation. For example, there are several interpretations one could draw from this analysis regarding the relationship between interest in programming and gender. The initial significant difference in interest between boys and girls could have been influenced by other factors not included in the study (e.g., pre-knowledge of programming, interests in designing programmable patterns, etc.). It is important to note that although the average score has decreased, this does not necessarily mean that the intervention has "failed." It could be that participants have a clearer understanding of programming after the intervention and can more accurately evaluate their interests (maybe they find it also too difficult or not fast forward). This could be more of an issue for boys than girls. It's possible that gender stereotypes about programming may have influenced the initial interest levels as well. For example, boys might have been more interested in T1 due to societal expectations or biases. These may have been less influential at T2, leading to a smaller difference in interest (note that the mean values were still between agree and strongly agree).

Especially in the area of self-efficacy, the average values unfortunately do not show a positive picture and indicate a clear deterioration. However, the analysis also showed that the questions were not optimally selected to cover this aspect. It can be concluded that the intervention had a differential effect on participants perceived self-efficacy during programming. The overall decrease in self-efficacy scores from T1 to T2 could possibly be due to the challenges and complexities encountered during the programming intervention. This could also be seen in the fact that many switched from their own patterns to predefined tutorials. Furthermore, as indicated in the evaluation of the field notes, this many participants can be attributed to the app still being quite difficult to use, requiring a lot of help or tutorials to create a design. In particular, the question "I can learn to program" may have reflected participants'

difficulty or slow progress in grasping programming concepts, thus negatively influencing their self-efficacy beliefs. Since there were no significant gender differences in changes in self-efficacy, the effect of the intervention seemed to be similar across genders. Nevertheless, it would be crucial to further explore potential gender-specific factors influencing these perceptions.

The results revolving around a sense of belonging indicate that boys are more likely to be able to place themselves in these roles and feel a sense of belonging to programming. This underscores the need for more interventions to promote girls' interest and affiliation in programming. For example, the programming course could benefit from including more real-world examples that are relevant to girls or creating more supportive and inclusive learning environments. It would be advisable not to base the unit's solely on programming but to give a broader picture of CS, as shown in Section 2.

Results regarding the "fun and games" item indicate significant gender differences in enjoyment of games and programming, with boys reporting higher levels of fun than girls. The intervention aimed at tailoring creative and more holistic activities to girls' interests could help increase their engagement and enjoyment in these areas.

Finally, girls showed a higher level of agreement with both post-questions, indicating a greater sense of pride in their designs (question 5b), which was a significant finding.

In summary, the results showed:

- Both genders showed a general interest in programming, with boys showing significantly more interest than girls.
- Both genders exhibited a general decrease in interest and self-efficacy related to programming after the intervention.
- There were significant differences in the sense of belonging and enjoyment of games, with boys scoring higher in both aspects.

These findings may indicate that the intervention, although generally considered valuable, may not have been fully responsive to the needs and interests of all participants. To empower girls through their programming and achieve better overall outcomes, it may be beneficial to expand the scope of future interventions beyond just programming patterns. Incorporating elements identified in the literature as positive, such as role models and clarity about IT careers, could provide a more comprehensive and effective approach to improving girls' experiences in STEM education. We also expect that better results can be achieved with the new version of the app (available since 2020).

Regarding the questionnaire itself, most scales were reliable, with the exception of the self-efficacy scale, which had slightly lower internal consistency. A renewed use of the questionnaire in this context will be possible within the Swiss project "Making at School." Overall, the correlations in Figure 7 suggest that to promote an engaging, effective, and enjoyable learning experience in a programming course for children, it is beneficial to build children's self-efficacy, ensure that they feel a sense of belonging, and incorporate elements of fun and play into the curriculum. Because these elements are interrelated (most of them show a strong positive correlation), improvements in one area are likely to have a positive impact on the others. However, as with all correlational analyses, causality cannot be definitively determined from these results, and further studies may be needed to better understand the specific causal relationships and their implications for children's programming.

Regarding children's final projects, we observe that 69% of the girls actualized their own designs as final products, while only 59% of the boys did so (see Figure 8). This discrepancy may indicate the girls' interest in bringing their own designs to life and boosting their motivation for programming. Girls frequently adhered to their initial designs, making adjustments or modifications based on available resources or time constraints, yet their unique ideas remained distantly identifiable.

Students encounter challenges in programming, such as converting a pattern into programming code using geometry. These issues are spread across various modalities, involving interactive identification of errors, solution development, and testing. A comparable problem was also observed in Study B by [36].

The girls who programmed with the help of tutorials often used the Ying-Yang symbol, swords, and crowns (see Figure 9), while the boys frequently chose the triangular spiral. This tutorial appeared to be a favorite among boys. Additionally, for their own designs, girls favored animals, closely followed by letters, names, and geometric shapes. On the other hand, boys tended to prefer letters, names, and geometric shapes such as smiley faces, hearts, and diamonds for their final products. Some girls created nature or food shapes, while none of the boys did. It is evident that girls and boys tend to select different shapes or designs for their final products. Therefore, it is crucial to present them with a wide range of design possibilities to spark their interest and enhance their motivation to create unique designs.

5.1 Lessons learned for the design process (design-workflow)

We were able to derive the following insights from the code and stitch units:

- There may be gender-specific differences in students' perceptions of intrinsic motivation when it comes to programming. There is room for improvement, especially in addressing gender disparities and measuring self-efficacy. It may be beneficial to integrate a more comprehensive understanding of CS and real-life applications to sustain students' engagement and motivation.
- The app used in the study was difficult to use, which may have contributed to the decline in self-efficacy and other items among students. It may be advisable to incorporate additional tutorials and resources to support students, especially in the basics of mathematics (circle, triangle, angles, etc.) and programming, particularly if they are doing so for the first time.
- Girls tend to be more interested in realizing their own designs as final products, while boys tend to choose existing shapes or designs for their final products. It may be helpful to expose students to a wide range of possible design options to spark their interest and increase their motivation.
- Embroidery machines and the results of self-designed work hold significant value for students, particularly for girls. This could enhance motivation and interest in technology and computer science.

Further lessons learned from the field notes:

- An important part of the process is the design. It is crucial to make it clear to the participants which patterns are easy and which are more difficult to implement

in programming code. It was beneficial that the participants emailed us their designs in advance so that we could think about different support options.
- Teachers should be involved so that they can conduct their own courses in the future.
- Initial programming of designs could be guided, for instance, through tutorials with CS teachers.
- After taking the initial steps, some students displayed great talent and found joy in programming independently.
- Students still wear their shirts a year after the course. They are very proud of their work and such as to show it off.

6 CONCLUSION AND OUTLOOK

The study conducted provided valuable insights into girls' and boys' motivation and interest in programming patterns. The first study question aimed to understand whether such activities promote intrinsic motivation, especially in girls. The second study question examined gender differences in the design and pattern-making process during stitching and coding lessons. In summary, our study underscores the importance of gender-sensitive pedagogical approaches in programming education. The findings highlight the need to develop interventions that promote a sense of belonging, enhance self-efficacy, and address the specific interests of boys and girls alike. The insights gained through this research are crucial to efforts to effectively engage all students in computer science.

In addition, the COVID-19 pandemic was an unexpected event that forced many projects to spontaneously rethink and move to online or hybrid teaching. In our case, this challenge led to some progress, as described in this section. After homeschooling started, the planned classroom sessions in cycle 2 could not be held in the originally planned form. However, we were able to quickly switch to hybrid course formats with increased resource use. Videos and online tutorials were created for students to access at any time. The concept of "bring your own device (BYOD)" was well promoted, especially in hybrid lessons, due to hygiene regulations, which alleviated teachers' concerns. Remote workshops and online coaching sessions for students were conducted by trainers and staff at TU Graz. The existing course concepts were expanded to include online formats. Course materials had to be adapted (see Figure 1). As a result of the project, a blended learning format is now available for code and stitch courses (also accessible on the Wiki).

Furthermore, on the well-known website hourofcode.org, which offers many tutorials for all age groups in more than 45 languages, a course based on the workshops could also be added. The website is used by millions of students and teachers in over 180 countries. Moreover, new stitches have been implemented. We have already succeeded in filling areas of objects with the app (see Figure 13). This type of stitch is called tatami-stitch. It is not about the stitch itself but more about the algorithm to optimally divide the area and then fill it with the stitch type. This functionality clearly sets the Embroidery Designer App apart from other comparable embroidery design programming environments. Furthermore, a connection with e-textiles, or "smart wearables," is interesting. Here, finished embroidered patterns can be extended, for example, with the help of sew-able LEDs or single-board computers such as BBC micro-bits (see Figure 14). Instructions have also been prepared for this on the Wiki.

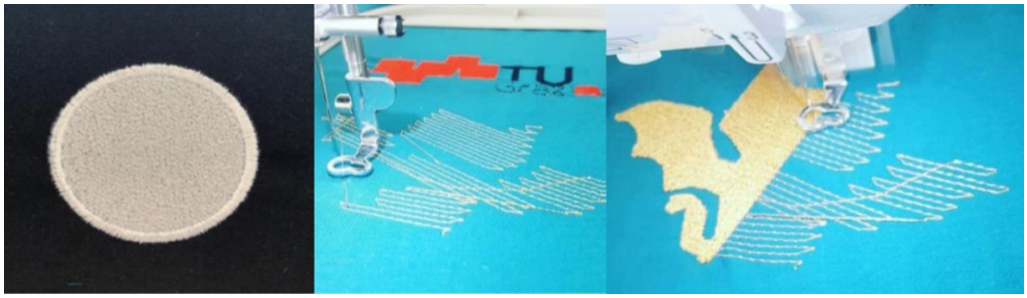

Fig. 13. Area filling of objects

Fig. 14. Connection of embroidery designs and e-textiles or to "smart wearables"

Findings, especially from cycle 2, have led to two new projects. Collaborating with bits4kids and TU Graz, an online course website was initiated and didactically prepared for the project "online platform" in 2021. This project was funded by the Austrian Research Promotion Agency (FFG). which also supported "Code'n'Stitch." In the end, children, both in and out of school, were able to utilize a learning management platform and a Discord server for support in workshops. bits4kids has conducted approximately 10 workshops in last 2 years, held both on-site in schools and during summer courses, as well as online through an "Online Coding Club" or during "Online Coding Weeks." The TU Graz team incorporated their findings into various workshops such as "Girls Coding Week" or "Maker Days for Kids." At Zurich University of Technology, the project results sparked a new initiative: "Making at School" (https://explore-making.ch [46]), with teacher training and workshops on digital designs, e-textiles, and "smart wearables." While the concept is still under refinement, the foundational idea from the funding project remains stable.

A final project conducted by bits4kids, which will start in 2022, aims to inspire girls with embroidery designs for computational thinking as a creative activity. The core element is the digital ecosystem with the network learning world, "KOALA." This initiative is tailored to meet the needs of girls.

- Learning independently and self-directed with creative, gender-appropriate tutorials.
- Learning through female coaching by women for women, also serving as role models.
- Learning through networking and exchange can take place among individuals in a secure virtual space.

Finally, in December 2022, the Embroidery Designer App was also awarded the Europe Social Impact Award by Huawei.

7 REFERENCES

[1] European Commission, "Study on the movement of skilled labour. Final report," *Publications Office of the European Union*, 2018. Available: https://op.europa.eu/en/publication-detail/-/publication/05079c0e-fc2e-11e8-a96d-01aa75ed71a1/ [Accessed: Jan. 20, 2024].

[2] World Economic Forum, "The future of jobs report," weforum.org, 2018. Available: https://www.weforum.org/reports/the-future-of-jobs-report-2018/ [Accessed: Feb. 11, 2024].

[3] European Statista, "Female students underrepresented in ICT," Eurostat, 2019. Available: https://ec.europa.eu/eurostat/en/web/products-eurostat-news/-/edn-20190425-1 [Accessed: Jan. 20, 2024].

[4] S. Gursch *et al.*, "Inequalities for women in science, technology and innovation," in *5th International Conference on Gender Research*, 2022, vol. 5, no. 2, pp. 88–96. https://doi.org/10.34190/icgr.5.1.157

[5] European Statista, "Share of female workers worldwide in 2022, by industry," Eurostat, 2022. Available: https://www.statista.com/statistics/1212105/share-o-female-workers-worldwide-by-professional-cluster/ [Accessed: Feb. 11, 2024].

[6] European Statista, "Share of women employed in science and technology in Central and Eastern Europe in 2022," Eurostat, 2022. Available: https://www.statista.com/statistics/1401935/cee-share-of-women-employed-in-science-and-technology/ [Accessed: Feb. 11, 2024].

[7] I. Kriesi and A. Basler, "The development of career aspirations of young women and men in Switzerland," *Social Change in Switzerland*, no. 23, 2020. https://doi.org/10.22019/SC-2020-00006

[8] B. Spieler, L. Oates-Indruchovà, and W. Slany, "Female students in computer science education: Understanding stereotypes, negative impacts, and positive motivation," *Journal of Women and Minorities in Science and Engineering*, vol. 26, no. 5, pp. 473–510, 2020. https://doi.org/10.1615/JWomenMinorScienEng.2020028567

[9] M. E. Beltŕan *et al.*, "Inclusive gaming creation by design in formal learning environments: 'Girly-girls' user group in no one left behind," in *International Conference of Design, User Experience, and Usability*, 2015, pp. 153–161. https://doi.org/10.1007/978-3-319-20898-5_15

[10] S. Gursch, "Coding initiative provides different approaches to inspire girls for programming," in *30th International Conference on Computers in Education Conference Proceedings*, vol. I, 2022, pp. 655–659.

[11] B. Spieler, V. Krnjic, W. Slany, K. Horneck, and U. Neudorfer, "Design, code, stitch, wear, and show it! mobile visual pattern design in school contexts," in *2020 IEEE Frontiers in Education Conference (FIE)*, 2020, pp. 1–9. https://doi.org/10.1109/FIE44824.2020.9274120

[12] B. Spieler, M. Grandl, and V. Krnjic, "The happy-lab: A gender-conscious way to learn coding basics in an open makerspace setting," in *Ceur Workshop Proceedings*, vol. 2755, 2020, pp. 64–75.

[13] B. Spieler and V. Krnjic, "Creative, engaging, and playful making-activities with smartphones and embroidery machines," in *Fablearn Europe/Makeed 2021-An International Conference on Computing, Design and Making in Education*, 2021, pp. 1–4. https://doi.org/10.1145/3466725.3466745

[14] B. Hesse, "Update for all – Teaching practice for fair digitalization," in B. Liedtke, B. Mauß, and M. Rother, Eds., University Press of TU Berlin, 2021.
[15] T. Luo, W. W. M. So, Z. H. Wan, and W. C. Li, "Stem stereotypes predict students' stem career interest via self-efficacy and outcome expectations," *International Journal of STEM Education*, vol. 8, no. 36, pp. 1–13, 2021. https://doi.org/10.1186/s40594-021-00295-y
[16] J. Wang, H. Hong, J. Ravitz, and M. Ivory, "Gender differences in factors influencing pursuit of computer science and related fields," in *Proceedings of the 2015 ACM Conference on Innovation and Technology in Computer Science Education*, 2015, pp. 117–122. https://doi.org/10.1145/2729094.2742611
[17] M. M. McGill, A. Decker, and A. Settle, "Undergraduate students' perceptions of the impact of pre-college computing activities on choices of major," *ACM Transactions on Computing Education (TOCE)*, vol. 16, no. 4, pp. 1–33, 2016. https://doi.org/10.1145/2920214
[18] S. Hidi and K. A. Renninger, "The four-phase model of interest development," *Educational Psychologist*, vol. 41, no. 2, pp. 111–127, 2006. https://doi.org/10.1207/s15326985ep4102_4
[19] L. Happe, B. Buhnova, A. Koziolek, and I. Wagner, "Effective measures to foster girls' interest in secondary computer science education," *Education and Information Technologies*, vol. 26, no. 3, pp. 2811–2829, 2021. https://doi.org/10.1007/s10639-020-10379-x
[20] A.-J. Lakanen and T. Kärkkäinen, "Identifying pathways to computer science: The long-term impact of short-term game programming outreach interventions," *ACM Transactions on Computing Education (TOCE)*, vol. 19, no. 3, pp. 1–30, 2019. https://doi.org/10.1145/3283070
[21] German Society for Computer Science, "Computer Science Education in Germany," informatik-monitor.de, 2023. Available: https://informatik-monitor.de/2022-23 [Accessed: Feb. 11, 2024].
[22] Swiss Curriculum 21. "Curriculum 21," lehrplan21.ch, 2019. Available: https://www.lehrplan21.ch/ [Accessed: Feb. 13, 2024].
[23] Federal Ministry of Education, Science and Research. (n.d.). Digital basic education. Available: https://www.bmbwf.gv.at/Themen/schule/zrp/dibi/dgb.html [Accessed: Feb. 11, 2024].
[24] N. Pinkard, S. Erete, C. K. Martin, and M. McKinney de Royston, "Digital youth divas: Exploring narrative-driven curriculum to spark middle school girls' interest in computational activities," *Journal of the Learning Sciences*, vol. 26, no. 3, pp. 477–516, 2017. https://doi.org/10.1080/10508406.2017.1307199
[25] T. Vrieler, A. Nylen, and A. Cajander, "Computer science club for girls and boys–a survey study on gender differences," *Computer Science Education*, vol. 31, no. 4, pp. 431–461, 2021. https://doi.org/10.1080/08993408.2020.1832412
[26] B. Annis and R. Nesbitt, *Results at the Top: Using Gender Intelligence to Create Break-Through Growth*. John Wiley & Sons, 2017.
[27] M. Csikszentmihalyi, *Flow: The Psychology of Optimal Experience*. New York: Harper Perennial, 1991.
[28] B. Buhnova and L. Happe, "Girl-friendly computer science classroom: Czechitas experience report," in *European Conference on Software Architecture (ECSA 2020)*, 2020, pp. 125–137. https://doi.org/10.1007/978-3-030-59155-7_10
[29] U. Wolz, M. Auschauer, and A. Mayr-Stalder, "Code crafting with turtlestitch" in *ACM Siggraph 2019 Studio*, 2019, pp. 1–2. https://doi.org/10.1145/3306306.3328009
[30] Y. Kafai, G. Jayathirtha, M. Shaw, and L. Morales-Navarro, "Codequilt: Designing an hour of code activity for creative and critical engagement with computing," in *IDC '21: Proceedings of the 20th Annual ACM Interaction Design and Children*, 2021, pp. 573–576. https://doi.org/10.1145/3459990.3465187

[31] L. Buechley, N. Elumeze, and M. Eisenberg, "Electronic/computational textiles and children's crafts," in *Proceedings of the 2006 Conference on Interaction Design and Children*, 2006, pp. 49–56. https://doi.org/10.1145/1139073.1139091

[32] A. Merkouris, K. Chorianopoulos, and A. Kameas, "Teaching programming in secondary education through embodied computing platforms: Robotics and wearables," *ACM Transactions on Computing Education (TOCE)*, vol. 17, no. 2, pp. 1–22, 2017. https://doi.org/10.1145/3025013

[33] A. C.-Y. Chen and Y.-C. Lin, "Girls in robot class: Smart textiles interactive tool-kits to enhance the participatory of women in technology," in *International Conference on Learning and Collaboration Technologies*, 2017, pp. 134–147. https://doi.org/10.1007/978-3-319-58515-4_11

[34] W. Slany, K. K. Luhana, M. Müller, C. Schindler, and B. Spieler, "Rock bottom, the world, the sky: Catrobat, an extremely large-scale and long-term visual coding project relying purely on smartphones," in *Constructionism 2018, Vilnius: Constructionism, Computational Thinking and Educational Innovation*, V. Dagienė and E. Jasutė (Hrsg.), 2018, pp. 104–119.

[35] L. Martin, "The promise of the maker movement for education," *Journal of Pre-College Engineering Education Research (J-PEER)*, vol. 5, no. 1, p. 4, 2015. https://doi.org/10.7771/2157-9288.1099

[36] K. A. Searle, B. K. Litts, and Y. B. Kafai, "Debugging open-ended designs: High school students' perceptions of failure and success in an electronic textiles design activity," *Thinking Skills and Creativity*, vol. 30, pp. 125–134, 2018. https://doi.org/10.1016/j.tsc.2018.03.004

[37] B. Spieler and S. Gursch, "Data survey codeNstitch," Zenodo, 2023. http://doi.org/10.5281/zenodo.7635652

[38] B. Spieler and S. Gursch, "Data designs codeNstitch," Zenodo, 2023. http://doi.org/10.5281/zenodo.7642304

[39] G. M. Sullivan and A. R. Artino, JR., "Analyzing and interpreting data from likert-type scales," *Journal of Graduate Medical Education*, vol. 5, no. 4, pp. 541–542, 2013. https://doi.org/10.4300/JGME-5-4-18

[40] R. Schwarzer and M. Jerusalem, "Generalized self-efficacy scale," APA PsycTests, 1995. https://doi.org/10.1037/t00393-000

[41] S. Krieger, M. Allen, and C. Rawn, "Are females disinclined to tinker in computer science?" in *Proceedings of the 46th ACM Technical Symposium on Computer Science Education*, 2015, pp. 102–107. https://doi.org/10.1145/2676723.2677296

[42] F. W. Li and C. Watson, "Game-based concept visualization for learning programming," in *Proceedings of the Third International ACM Workshop on Multimedia Technologies for Distance Learning*, 2011, pp. 37–42. https://doi.org/10.1145/2072598.2072607

[43] B. Tanujaya, R. C. I. Prahmana, and J. Mumu, "Likert scale in social sciences research: Problems and difficulties," *FWU Journal of Social Sciences*, vol. 16, no. 4, pp. 89–101, 2022. http://ojs.sbbwu.edu.pk/fwu-journal/index.php/ojss/article/view/871

[44] P. A. Mabe and S. G. West, "Validity of self-evaluation of ability: A review and meta-analysis," *Journal of Applied Psychology*, vol. 67, no. 3, pp. 280–296, 1982. https://doi.org/10.1037//0021-9010.67.3.280

[45] J. Saldana, *The Coding Manual for Qualitative Researchers*. Thousand Oaks, CA: Sage Publications, 2021.

[46] B. Spieler, T. M. Schifferle, and M. Dahinden, "Making Im Unterricht: Erfahrungen Von Lehrpersonen Aus Dem Design-Based-Research-Projekt," *MedienPädagogik: Zeitschrift für Theorie Und Praxis Der Medienbildung 56 (Making & more)*, pp. 331–363, 2024. https://doi.org/10.21240/mpaed/56/2024.02.14.X

8 AUTHORS

Bernadette Spieler is a Professor of Computing Skills in Education at Zurich University of Teacher Education, Switzerland. She received her Ph.D. from Graz University of Technology, Institute of Software Engineering in 2018. Her research interests include fundamental computer science education, AI didactics, e-learning, maker education, programming concepts, game design and software development. Spieler is particularly dedicated to developing engaging computer science activities for children and teenagers, aiming to enhance the learning experiences of girls in particular (E-mail: bernadette.spieler@phzh.ch).

Sarina Gursch is with the Graz University of Technology, Graz, Austria.

Vesna Krnjic is with the Catrobat Association, Graz, Austria.

Karin Horneck is with the bits4kids OG, Graz, Austria.

Wolfgang Slany is with the Catrobat Association, Graz, Austria; Graz University of Technology, Graz, Austria.

Imprint

iJET – International Journal of Emerging Technologies in Learning
http://www.i-jet.org

Indexing
International Journal of Emerging Technologies in Learning is indexed in DBLP, Ulrich, EBSCO, and LearnTechLib.

Publication Frequency
Bi-Monthly

ISSN
1863-0383

Publisher
International Federation of Engineering Education Societies (IFEES)
IFEES-GEDC Secretariat
c/o George Mason University
Volgenau School of Engineering
4400 University Drive, MS 4A3
Fairfax, VA 22030
USA

Editorial Office
CTI Global
Frankfurt, Vienna, New York, Bengaluru, Hong Kong
Madrider Strasse 4
60327 Frankfurt am Main
Germany