Haoran Xie · Chiu-Lin Lai ·
Wei Chen · Guandong Xu ·
Elvira Popescu (Eds.)

LNCS 14409

Advances in Web-Based Learning – ICWL 2023

22nd International Conference, ICWL 2023
Sydney, NSW, Australia, November 26–28, 2023
Proceedings

Springer

Lecture Notes in Computer Science 14409

The series Lecture Notes in Computer Science (LNCS), including its subseries Lecture Notes in Artificial Intelligence (LNAI) and Lecture Notes in Bioinformatics (LNBI), has established itself as a medium for the publication of new developments in computer science and information technology research, teaching, and education.

LNCS enjoys close cooperation with the computer science R & D community, the series counts many renowned academics among its volume editors and paper authors, and collaborates with prestigious societies. Its mission is to serve this international community by providing an invaluable service, mainly focused on the publication of conference and workshop proceedings and postproceedings. LNCS commenced publication in 1973.

Haoran Xie · Chiu-Lin Lai · Wei Chen ·
Guandong Xu · Elvira Popescu
Editors

Advances in Web-Based Learning – ICWL 2023

22nd International Conference, ICWL 2023
Sydney, NSW, Australia, November 26–28, 2023
Proceedings

Springer

Editors
Haoran Xie
Lingnan University
Hong Kong, Hong Kong

Chiu-Lin Lai
National Taipei University of Education
Taipei, Taiwan

Wei Chen
Chinese Academy of Agricultural Sciences
Beijing, China

Guandong Xu
University of Technology Sydney
Sydney, NSW, Australia

Elvira Popescu
University of Craiova
Craiova, Romania

ISSN 0302-9743 ISSN 1611-3349 (electronic)
Lecture Notes in Computer Science
ISBN 978-981-99-8384-1 ISBN 978-981-99-8385-8 (eBook)
https://doi.org/10.1007/978-981-99-8385-8

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd.
The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

Preface

This volume presents the contributions to the 22nd edition of the annual International Conference on Web-based Learning (ICWL). The first edition of ICWL was held in Hong Kong in 2002. Since then, it has been held 21 more times, on three continents: Melbourne, Australia (2003); Beijing, China (2004); Hong Kong, China (2005, 2011); Penang, Malaysia (2006); Edinburgh, UK (2007); Jinhua, China (2008); Aachen, Germany (2009); Shanghai, China (2010); Sinaia, Romania (2012); Kenting, Taiwan (2013); Tallinn, Estonia (2014); Guangzhou, China (2015); Rome, Italy (2016); Cape Town, South Africa (2017); Chiang Mai, Thailand (2018); Magdeburg, Germany (2019); Ningbo, China (2020); Macao, China (2021); Tenerife, Spain (2022). This year, ICWL 2023 was held on 26–28 November, 2023 in Sydney, Australia, organized by University of Technology Sydney, Australia.

The topics proposed in the ICWL Call for Papers included several relevant issues, ranging from Semantic Web for E-Learning, through Learning Analytics, Computer-Supported Collaborative Learning, Assessment, Pedagogical Issues, E-learning Platforms, and Tools, to Mobile Learning.

We received 23 submitted contributions. All of the submitted papers were assigned to three members of the Program Committee (PC) for peer review. Reviews were single-blind, and on average submissions received 3.4 reviews. All reviews were checked and discussed by the PC chairs, and additional reviews or meta-reviews were elicited if necessary.

Finally, we accepted 9 full papers, with an acceptance rate of 39%. We also included additional contributions as short papers (7) in the proceedings.

Furthermore, the conference continued the traditional initiative, started by ICWL 2016, of holding the 8th International Symposium on Emerging Technologies for Education (SETE) at the same location. SETE collected the traditional workshop activities managed by ICWL in the past years and additionally featured an organization in tracks. Workshops and tracks added new and hot topics on technology-enhanced learning, providing a better overall conference experience to the ICWL and SETE attendees.

ICWL 2023 was made possible by the joint efforts of many people and institutions. There is a long list of people who volunteered their time and energy to put together the conference and who deserve special thanks.

We would like to thank all the PC members for their great effort in reading, commenting on, and finally selecting the papers. We also thank all the external reviewers for assisting the PC in their particular areas of expertise.

We sincerely thank the authors of all submitted papers and all the conference attendees. Thanks are also due to the staff at Springer for their help with producing the proceedings and to the developers and maintainers of the EasyChair software, which greatly helped simplify the submission and review process.

September 2023

Haoran Xie
Chiu-Lin Lai
Wei Chen
Guandong Xu
Elvira Popescu

Organization

General Co-chairs

Guandong Xu	University of Technology Sydney, Australia
Elvira Popescu	University of Craiova, Romania

Program Committee Co-chairs

Haoran Xie	Lingnan University, Hong Kong SAR
Chiu-Lin Lai	National Taipei University of Education, Taiwan
Wei Chen	Agricultural Information Institute of CAAS, China

Steering Committee Representatives

Qing Li	Hong Kong Polytechnic University, Hong Kong SAR
Rynson W. H. Lau	City University of Hong Kong, Hong Kong SAR

Workshop Co-chairs

Luigi Laura	International Telematic University Uninettuno, Italy
Robin Braun	University of Technology Sydney, Australia
Nana Jin	Shenzhen University, China

Publicity Co-chairs

Ivana Marenzi	Leibniz University of Hanover, Germany
Minjie Hu	Wellington Institute of Technology, New Zealand
Junjie Wu	Shenzhen Technology University, China

Local Organization Co-chairs

Huan Huo	University of Technology Sydney, Australia
Md Rafiqul Islam	Australian Institute of Higher Education, Australia

Web Chair

Kaize Shi	University of Technology Sydney, Australia

Program Committee

Silvana Aciar	Universidad Nacional de San Juan, Argentina
Manuel Caeiro Rodríguez	University of Vigo, Spain
Cristian Cechinel	Federal University of Pelotas, Brazil
Chih-Ming Chen	National Chengchi University, Taiwan
Jia Chen	Central China Normal University, China
Wei Chen	Agricultural Information Institute, Chinese Academy of Agricultural Sciences, China
Shu-Chen Cheng	Southern Taiwan University of Science and Technology, Taiwan
Dickson K. W. Chiu	University of Hong Kong, Hong Kong SAR
Maria-Iuliana Dascalu	University POLITEHNICA of Bucharest, Romania
Damiano Distante	University of Rome Unitelma Sapienza, Italy
Shi Dong	Wuhan University, China
Ramon Fabregat	Universitat de Girona, Spain
Baltasar Fernandez-Manjon	Universidad Complutense de Madrid, Spain
Manuel Freire Morán	Universidad Complutense de Madrid, Spain
Alicia García-Holgado	Universidad de Salamanca, Spain
Denis Gillet	Swiss Federal Institute of Technology in Lausanne (EPFL), Switzerland
Carina Gonzalez-González	Universidad de La Laguna, Spain
Sabine Graf	Athabasca University, Canada
Tianyong Hao	South China Normal University, China
Michael A. Herzog	Magdeburg-Stendal University, Germany
Chuanhui Huang	South-Central Minzu University, China
Mirjana Ivanovic	University of Novi Sad, Serbia
Tomaž Klobučar	Jožef Stefan Institute, Slovenia
Milos Kravcik	DFKI GmbH, Germany

Amruth Kumar	Ramapo College of New Jersey, USA
Chiu-Lin Lai	National Taipei University of Education, Taiwan
Luigi Laura	International Telematic University Uninettuno, Italy
Frederick Li	University of Durham, UK
Carla Limongelli	Università Roma Tre, Italy
Koong Lin	National University of Tainan, Taiwan
An Liu	Soochow University, China
Zhi Liu	Central China Normal University/National Engineering Research Center for E-Learning, China
Martín Llamas-Nistal	Universidade de Vigo, Spain
Jia-Jiunn Lo	Chung-Hua University, Taiwan
Matteo Lombardi	Griffith University, Australia
Ivana Marenzi	L3S Research Center, Germany
Fernando Moreira	Universidade Portucalense, Portugal
Wolfgang Mueller	University of Education Weingarten, Germany
Claudia Pons	Universidad Nacional de La Plata, Argentina
Elvira Popescu	University of Craiova, Romania
Francesca Pozzi	Istituto Tecnologie Didattiche - CNR, Italy
Demetrios Sampson	Curtin University, Australia
Filippo Sciarrone	Universitas Mercatorum, Italy
Lei Shi	Newcastle University, UK
Marcus Specht	Delft University of Technology, The Netherlands
Yuan Tian	Central China Normal University, China
Carlos Vaz de Carvalho	Instituto Superior de Engenharia do Porto, Portugal
Pierpaolo Vittorini	University of L'Aquila, Italy
Riina Vuorikari	Institute for Prospective Technological Studies, Spain
Haipeng Wan	Capital Normal University, China
Tao Wang	Central China Normal University, China
Carine Webber	Universidade de Caxias do Sul, Brazil
Shiting Wen	Zhejiang University, China
Tak-Lam Wong	Douglas College, Canada
Charoenchai Wongwatkit	Mae Fah Luang University, Thailand
Pengfei Wu	Beijing Normal University, China
Sheng-Yi Wu	National Pingtung University, Taiwan
Ting-Ting Wu	National Yunlin University of Science and Technology, China
Dan Xia	Central China Normal University, China
Haoran Xie	Lingnan University, Hong Kong SAR

Guandong Xu	University of Technology Sydney, Australia
Zhenguo Xu	Qufu Normal University, China
Qi Zhang	Huaibei Normal University, China
Nan Zhou	Capital Normal University, China
Vincent Y. Zhuang	Hong Kong Polytechnic University, Hong Kong SAR
Francisco Javier Álvarez Rodríguez	Universidad Autónoma de Aguascalientes, Mexico

Additional Reviewers

Anwar, Muhammad
Li, Zhaoxing
Zhang, Ruofei
Zhong, Yingyu
Zhu, Qile

Contents

Learning Content Management

Computer Support for Intelligent Tutoring

Massive Open Online Courses and Learning Analytics

Intention of MOOCs Adoption, Completion and Continued Use

Mansor Alzahrani(✉), Faezeh Karimi, Gnana Bharathy, and Mukesh Prasad

University of Technology Sydney, 15 Broadway, Ultimo, NSW 2007, Australia
mansor.alzahrani@student.uts.edu.au, {faezeh.karimi, gnana.bharathy,mukesh.prasad}@uts.edu.au

Abstract. MOOCs were introduced almost a decade ago, providing knowledge to learners without financial support. Despite the advantages, low completion rates remain a persistent issue. This study focuses solely on individual adoption of MOOCs, seeking to understand why people adopt them in the first place, what motivates them to complete their courses and whether they continue learning through MOOCs after completing them. The aim of the current work is to investigate individuals' intention towards the three stages of adoption: the initial adoption of MOOCs, the completion of MOOCs they started and the continued use of MOOCs beyond the completion of the subject they enrolled at.

Keywords: Adoption · MOOCs · Behavioral intention · TUE

1 Introduction

MOOCs were introduced almost a decade ago, providing knowledge to learners without financial support. Despite the advantages, low completion rates remain a persistent issue (Jordan 2015; Sallam 2017). The adoption of technology studies possible variables that influence adoption, including behaviour and intention (Ajzen 1991). Various theories and models examine technology adoption at individual and organisational levels (Taherdoost 2018; Khan and Qudrat-Ullah 2021). This study focuses solely on individual adoption of MOOCs, seeking to understand why people adopt them in the first place (Abu-Shanab and Musleh 2018; Al-Adwan 2020; Haron et al. 2021), what motivates them to complete their courses (Hone and El Said, 2016; Wang and Baker 2018; Romero-Rodríguez et al. 2020), and whether they continue learning through MOOCs after completing them (Alraimi, Zo and Ciganek 2015; Wu and Chen 2017). This study aims to investigate individuals' intention towards the three stages of adoption: the initial adoption of MOOCs, the completion of MOOCs they started and the continued use of MOOCs beyond the completion of the subject they enrolled at.

2 Theoretical Background and Research Model

Research has extensively investigated the behavioural intention towards technology use (Shah, Khanna and Patel 2021), revealing the significant influence of intention on future behaviour (Ajzen 1991). Researchers have utilised models like TAM and UTAUT

H. Xie et al. (Eds.): ICWL 2023, LNCS 14409, pp. 3–12, 2023.
https://doi.org/10.1007/978-981-99-8385-8_1

to explore how various factors relate to the adoption and intention to continue using MOOCs. However, some researchers have opted for a more context-focused and comprehensive approach, including Ma and Lee (2019), who employed the Technology-User-Environment (TUE) model. Their model was adapted from the well-established Technology-Organisation-Environment (TOE) model by Tornatzky et al. (1990). TUE focuses on individual rather than organisational adoption. In spite of the fact that TUE is relatively new model, it is a promising approach to understanding technology adoption behaviour. It covers three different contexts: technology, user, and social-environmental. Numerous studies were conducted to separately investigate the behavioural intention for the initial adoption, completion, and continued use of MOOCs. However, only a limited number of studies have explored all of the three aspects together. Although completion and continued use may sound similar, they are not. Completion intention refers to a student intention to finish the course while continuance intention refers to their willingness to keep learning beyond the completion of a course (Gupta and Maurya 2020). Therefore, this work aims to fill that gap by investigating the three intentions (adoption, completion, and continued use beyond completion) and identifying which factors are most relevant and influential at each stage. To address this, the current research will expand the TUE model by adding a new context, the Learning Context, and investigate the underlying constructs of all four contexts and their relationships with MOOCs adoption, completion, and continued use.

2.1 Conceptual Model

As mentioned earlier, the current work will adopt the TUE model to conduct the research. The model has been modified to suit the purpose of the study. Five constructs were added to the original model and presented in green colour (as shown in Fig. 1).

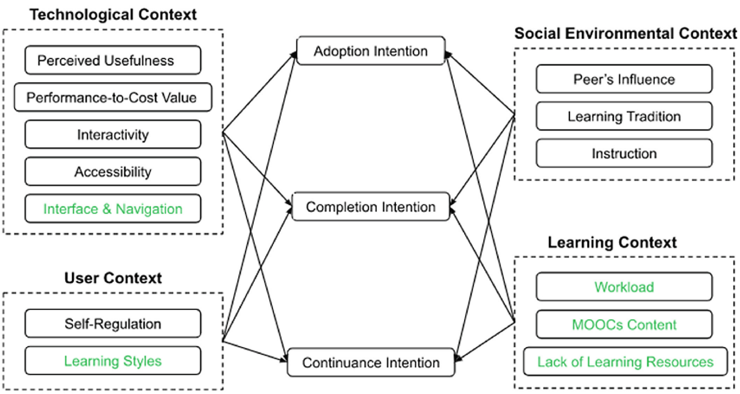


Fig. 1 The TUE Model adapted by Ma & Lee (2019) and modified by authors.

Technological Context. This context contains the technological factors that may influence the behavioural intention of an individual towards any of the three actions: to adopt, complete and/or continue using MOOCs. This context has five constructs perceived usefulness, performance-to-cost value, accessibility, interactivity and interface and navigation.

Perceived Usefulness. Perceived usefulness (PU) is a critical factor in technology adoption, defined as "the degree to which a user believes that using a particular system would enhance his or her job performance" (Davis 1989). PU has been identified as a key determinant of the intention to use MOOCs (Chu et al. 2015) and continue using them (Wu and Chen 2017). Ma and Lee (2020) define perceived usefulness in the context of MOOCs as "the degree to which a learner believes learning [through] MOOCs would enhance his or her academic or job performance". Therefore, it is hypothesised that:

- H1: Perceived usefulness influences individuals' intention to adopt, complete and continue using MOOCs.

Performance-to-Cost Value. Venkatesh et al. (2012) note that cost influences individuals' intention to use technology. Adoption of technology depends on a cost-benefit analysis, and knowledge seekers are more likely to use MOOCs if they believe it offers acceptable value for the cost (Davis 1989). Performance-to-cost value is defined as the degree to which learners believe using MOOCs is worthwhile in terms of money, time, and effort invested (Ma and Lee 2020). Additionally, Zeithaml (1988) suggests that individuals are more likely to adopt an innovation if they perceive it will improve their overall utility, which can include time and effort as well. Therefore, MOOCs adoption may depend on learners' perception of the value they will receive in return for their investment of time and effort. Therefore, it is hypothesised that:

- H2: Performance-to-cost value influences individuals' intention to adopt, complete and continue using MOOCs.

Interactivity. Interaction is crucial in the learning process. Some learning theories, such as the Cooperative Learning theory (Gillies 2016), suggest that learning efficiency can be improved through group work, which necessitates social interaction among students. Moore's (1989) typology of interactions, including learner-content, learner-learner, and learner-instructor, highlights the importance of incorporating all three types of interaction in online learning programs. However, some MOOC platforms lack interaction channels, leading to demotivation and lower completion rates (Hone and El Said 2016). Conversely, learners who engage in forums tend to perform better (Gillani and Eynon 2014). Therefore, it is hypothesised that:

- H3: Lack of interactivity negatively influences individuals' intention to adopt, complete and continue using MOOCs.

Accessibility. MOOC accessibility is crucial for ensuring that they can be used effectively. However, various barriers, such as limited technology or internet access, language barriers, and high costs of quality courses, hinder learners' access to MOOCs (Sanchez-Gordon and Luján-Mora 2018). This is particularly relevant in developing countries, where education and technology are limited (Abdel-Maksoud 2019). Therefore, further research on the accessibility of quality MOOCs to those who need them should be conducted (Liyanagunawardena et al. 2014), and, thus, it is hypothesised that:

- H4: Accessibility influences individuals' intention to adopt, complete and continue using MOOCs.

Interface and Navigation. Designing MOOCs platforms can be vital to students' experience, engagement and knowledge (Gamage et al. 2020). Studies show that interface design and navigation affect students' intention to continue using MOOCs (Cheng 2021), with culture influencing localised MOOCs' web design (Liu et al. 2020). It is crucial to consider cultural differences when designing MOOCs to ensure that they are effective and meet students' needs. Therefore, it is hypothesised that:

- H5: Interface and navigation influence individuals' intention to adopt, complete and continue using MOOCs.

User Context. To comprehend the technology adoption process, it is essential to investigate individuals' personal traits and experiences that shape their decision-making as some research suggest that personality does influence intention of adoption (Svendsen et al. 2013). Examining factors within the individual context is necessary to understand technology adoption. This context covers self-regulation and learning styles constructs.

Self-regulation. Self-regulation skills are crucial for successful learning in online environments, such as MOOCs, where learners are responsible for their own learning (Rosé and Ferschke 2016). Cultural differences affect self-regulated learning performance (Tang 2021). It is essential to explore the relationship between self-regulation and MOOCs adoption, completion, and continued use among the multicultural MOOC audience. While some studies indicate that self-regulation has no impact on MOOC adoption (Gupta 2020), others suggest it can reduce dropout rates (Al-Adwan 2020), highlighting the need for additional research. Therefore, it is hypothesised that:

- H6: Self-regulation influences individuals' intention to adopt, complete and continue using MOOCs.

Learning Styles. Individuals have their preferred way of learning (Fleming and Mills 1992), commonly referred to as their learning style. Common learning styles include visual, auditory, reading, and kinesthetic, and students may benefit from a combination of these styles (Fleming 1995). Some students prefer one style, while others benefit from a combination of two or more (Urval et al. 2014). Neglecting learning styles in online learning can hinder student engagement and success (Idrizi et al. 2018). Supporting various learning styles is essential in MOOCs where interactions with instructors are limited (Corrado et al. 2021). It is hypothesised that:

- H7: Learning styles influence individuals' intention to adopt, complete and continue using MOOCs.

Social-Environmental Context. This section will discuss the social and environmental factors that may have influence on individuals' intentions to adopt, complete and/or continue using MOOCs. Peer's influence, learning tradition and instruction are three factors covered by the social-environmental context.

Peer's Influence. People tend to emulate others, primarily when it benefits them. The Social Contagion theory explains how social networks, including peers, can influence behaviour (Burgess et al. 2018). This theory also applies to technology adoption, as social contagion can increase demand for certain products (Langley et al. 2012). Peer influence can motivate individuals to adopt MOOCs (Tseng et al. 2019). It can be defined as "the degree to which a learner's decision to adopt MOOCs is influenced by his or her peers' attitudes and actions" (Ma and Lee 2020). This influence can extend to encompass completing enrolled MOOCs and continuing to enrol in others. Therefore, it is hypothesised that:

- H8: Peer's influences individuals' intention to adopt, complete and continue using MOOCs.

Learning Tradition. Traditional education involves face-to-face learning with a teacher, making it difficult for some students to adapt to online learning. Traditions, "an inherited body of customs and beliefs" (Handler and Linnekin 1984), are resistant to innovation, posing a challenge for MOOCs (Al-Adwan 2020), which require self-directed learning. Learning tradition refers to the conflict between traditional routines and MOOC learning (Ma and Lee 2020). Therefore, it is hypothesised that:

- H9: Learning tradition influences individuals' intention to adopt, complete and continue using MOOCs.

Instruction. Instruction is vital for learners to gain knowledge and skills. Studies show that proper instruction improves student performance (Krashen et al. 1978), and individualized instruction may be necessary for some learners to succeed (Fuchs et al. 2014). However, online learning platforms like MOOCs may lack sufficient instruction (Gardner and Brooks 2018), leading to dropouts (Hone & El Said, 2016). Instruction on MOOCs refers to the guidance and support provided by instructors on how to use the platform, select courses, and receive assistance when needed (Ma and Lee 2020). Therefore, it is hypothesised that:

- H10: Instruction influences individuals' intention to adopt, complete and continue using MOOCs.

Learning Context. Because of its importance and the difference of MOOCs purpose as a technology, *learning* has been added to the adopted model as a new separate context which covers different aspects that related to learning in general and particularly on MOOCs. The learning context contains a few constructs workload, MOOCs content, and lack of learning resources.

Workload. The workload required to pass a MOOC can affect a student's decision to enrol. Chang et al. (2015) found that about 58% of students drop out due to time management issues. However, the workload on MOOCs can also be intensive, where some studies found that 24% of students not completing learning activities due to their heaviness (Liu et al. 2014). Defining and categorizing workload intensity is necessary, as workload can be subjective and determined by various factors. For example, how demanding and the depth of the topic can affect workload. Further research, therefore, is required to differentiate between time management and workload. It is proposed that:

- H11: Workload influences individuals' intention to adopt, complete and continue using MOOCs.

MOOCs Content. MOOCs' have been criticised for long that MOOCs are not following the criteria of what can be called good learning, questioning its quality and efficiency (Conole 2013). Content quality is key to MOOCs' success (Hone and El Said 2016), which should feature a well-organized and interactive structure, good breadth and depth of topics covered, and up-to-date materials. Defining such features is crucial to ensure MOOCs' success (Deshpande and Chukhlomin 2017). Feedback from MOOC users is needed to assess content features affecting adoption, completion, and continued usage. Thus, it is proposed that:

- H12: MOOCs content influences individuals' intention to adopt, complete and continue using MOOCs.

Lack of Learning Resources. Limited resources, including finances, infrastructure, and internet access, have been shown to hinder MOOC adoption. This issue is prevalent in developing countries, including the Arab region (Adham and Lundqvist 2015; Pasha et al. 2016). In some parts of the world, a shortage of learning resources, such as quality content and experienced teachers, has also been identified (Ghavifekr et al. 2006). This research investigates the impact of limited learning resources on MOOC adoption, completion, and continued usage. Therefore, it is proposed that:

- H13: Lack of learning resources positively influences individuals' intention to adopt, complete and continue using MOOCs.

3 Upcoming Work

3.1 Study Design

MOOCs users are diverse, making it challenging to determine factors that influence their adoption, completion, and continued use of MOOCs. Therefore, it is important for this study to include a diverse sample to investigate these behaviours. To investigate influential factors on MOOC adoption, completion and continuation, it's vital to examine various MOOC platforms and their potential users, given differences in features and individual preferences. This quantitative study uses a survey questionnaire to cover various aspects of MOOCs and learning. Participants were requested to select one MOOC they have already taken or they were currently taking at the time of the survey to ensure that all answers are based on the same MOOC and platform.

3.2 Sampling and Measurements

The present study faces both easy and difficult aspects in examining factors that influence MOOCs adoption, completion, and continued use. It is easy due to the large number of people who are already MOOCs users or familiar with them, but difficult due to the diversity of MOOCs deliveries and features. To address this, the study collected around 1000 samples from people of different backgrounds, including those who have never heard of MOOCs, those who have heard of MOOCs but have not studied through MOOCs platforms, those who have failed to complete their MOOCs courses, and those who have completed their courses on MOOCs platforms. The data was collected through different social media platforms as well as through quality data collection platforms, namely CloudResearch and Profliric (Douglas et al. 2023). The survey for this study includes at least 3–4 questions for every construct, which may make it long and less likely to be completed by participants. However, having more questions in the questionnaire can lead to more precise inferences from the results, higher quality responses, and a more comprehensive assessment of the topic being studied (Steyn 2017). Most of the survey items were adopted from previous research, while new constructs and items were developed by the researcher. Smart-PLS tested the survey for validity and reliability, and the results showed that the survey questionnaire is valid and reliable.

4 Conclusion

The purpose of this paper was to examine how students' intentions to use massive open online courses (MOOCs), complete them, and continue using them after completion would be affected by various factors such as interactivity, learning styles, workload, lack of learning resources, and peer influence. Because of the lack of studies that bring the three together in one study, it was important to attempts to fill this gap by studying a broader audience, those who are familiar and unfamiliar with MOOCs, completers and non-completers. This will enhance our understanding of the persistent problem of low completion rates in MOOCs that has been ongoing for over a decade, leading to higher completion rates and more learners completing their MOOCs.

References

A. Khan, R., Qudrat-Ullah, H.: Technology adoption theories and models. In: Adoption of LMS in Higher Educational Institutions of the Middle East. ASTI, pp. 27–48. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-50112-9_5

Abdel-Maksoud, N.F.: Factors affecting MOOCs' adoption in the arab world: exploring learners' perceptions on MOOCs' drivers and barriers. Int. Educ. Stud. **12**(11), 164 (2019). https://doi.org/10.5539/ies.v12n11p164

Abu-Shanab, E., Musleh, S.: The adoption of massive open online courses: challenges and benefits. Int. J. Web-Based Learn. Teach. Technol. **13**(4), 62–76 (2018). https://doi.org/10.4018/IJWLTT.2018100104

Al-Adwan, A.S.: Investigating the drivers and barriers to MOOCs adoption: the perspective of TAM. Educ. Inf. Technol. **25**(6), 5771–5795 (2020). https://doi.org/10.1007/s10639-020-10250-z

Alraimi, K.M., Zo, H., Ciganek, A.P.: Understanding the MOOCs continuance: the role of openness and reputation. Comput. Educ. **80**, 28–38 (2015). https://doi.org/10.1016/j.compedu.2014.08.006

Ajzen, I.: The theory of planned behavior. Organ. Behav. Human Decis, Process. **50**(2), 179–211 (1991). https://doi.org/10.1016/0749-5978(91)90020-T

Burgess, L.G., Riddell, P.M., Fancourt, A., Murayama, K.: The influence of social contagion within education: a motivational perspective. Mind Brain Educ. **12**(4), 164–174 (2018). https://doi.org/10.1111/mbe.12178

Chang, R.I., Hung, Y.H., Lin, C.F.: Survey of learning experiences and influence of learning style preferences on user intentions regarding MOOCs. Brit. J. Educ. Technol. **46**(3), 528–541 (2015). https://doi.org/10.1111/bjet.12275

Cheng, Y.M.: Can gamification and interface design aesthetics lead to MOOCs' success? Educ. Train. **63**(9), 1346–1375 (2021). https://doi.org/10.1108/ET-09-2020-0278

Conole, G.: MOOCs as disruptive technologies: strategies for enhancing the learner experience and quality of MOOCs. Revista de Educación a Distancia (RED) **50** (2013). https://doi.org/10.6018/red/50/2

Corrado, R., Pretorius, E., Van Der Westhuizen, G.: Undergraduate students' experiences of the use of moocs for learning at a Cambodian University. Educ. Sci. **11**(7), 1–16 (2021). https://doi.org/10.3390/educsci11070336

Chu, R., Ma, E., Feng, Y., Lai, I. K.: Understanding learners' intention toward massive open online courses. In: Hybrid Learning: Innovation in Educational Practices: 8th International Conference, ICHL 2015, Wuhan, China, July 27-29, 2015, Proceedings 8, pp. 302–312. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-20621-9_25

Davis, F.D.: Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q. Manag. Inf. Syst. **13**(3), 319–339 (1989). https://doi.org/10.2307/249008

Deshpande, A., Chukhlomin, V.: What makes a good MOOC: a field study of factors impacting student motivation to learn. Am. J. Dist. Educ. **31**(4), 275–293 (2017). https://doi.org/10.1080/08923647.2017.1377513

Douglas, B.D., Ewell, P.J., Brauer, M.: Data quality in online human-subjects research: comparisons between MTurk, prolific, CloudResearch, Qualtrics, and SONA. Plos One **18**(3), e0279720 (2023). https://doi.org/10.1371/journal.pone.0279720

Fleming, N.D., Mills, C.: Not another inventory, rather a catalyst for reflection. Improve Acad. **11**, 20210331 (1992). https://doi.org/10.3998/tia.17063888.0011.014

Fleming, N.D.: I'm different; not dumb: modes of presentation (VARK) in the tertiary classroom. In: Research and Development in Higher Education, Proceedings of the Annual Conference of the Higher Education and Research Development Society of Australasia, vol. 18 (1995)

Fuchs, D., Fuchs, L.S., Vaughn, S.: Data based individualization: what is intensive instruction and why is it important? Teach. Except. Child. **46**(4), 13–18 (2014). http://ceedar.education.ufl.edu/mtss-udl-di-dev/pdfs-worksheets/mtss/Fuchs

Gamage, D., Perera, I., Fernando, S.: MOOCs lack interactivity and collaborativeness: evaluating MOOC platforms. Int. J. Eng. Pedagogy **10**(2), 94–111 (2020). https://doi.org/10.3991/ijep.v10i2.11886

Gardner, J., Brooks, C.: Student success prediction in MOOCs. User Model. User-Adapt. Interact. **28**(2), 127–203 (2018). https://doi.org/10.1007/s11257-018-9203-z

Ghavifekr, S., Kunjappan, T., Ramasamy, L., Anthony, A., My, E.: Teaching and learning with ICT tools: issues and challenges from teachers' perceptions. Malay. Online J. Educ. Technol. **4**(2), 38–57 (2006). www.mojet.net

Gillani, N., Eynon, R.: Communication patterns in massively open online courses. Internet High. Educ. **23**, 18–26 (2014). https://doi.org/10.1016/J.IHEDUC.2014.05.004

Gillies, R.M.: Cooperative learning: review of research and practice. Aust. J. Teach. Educ. **41**(3), 39–54 (2016). https://doi.org/10.14221/ajte.2016v41n3.3

Gupta, K.P.: Investigating the adoption of MOOCs in a developing country: application of technology-user-environment framework and self-determination theory. Interact. Technol. Smart Educ. **17**(4), 355–375 (2020). https://doi.org/10.1108/ITSE-06-2019-0033

Gupta, K.P., Maurya, H.: Adoption, completion and continuance of MOOCs: a longitudinal study of students' behavioural intentions. Behav. Inf. Technol. **41**(3), 611–628 (2022). https://doi.org/10.1080/0144929X.2020.1829054

Handler, R., and Linnekin, J. 1984. "Tradition, Genuine or Spurious," The Journal of American Folklore (97:385), University of Illinois Press, p. 273. (https://doi.org/10.2307/540610)

Haron, H., Hussin, S., Yusof, A.R.M., Samad, H., Yusof, H.: Implementation of the UTAUT model to understand the technology adoption of MOOC at public universities. In: IOP Conference Series: Materials Science and Engineering, vol. 1062, no. 1 (2021). https://doi.org/10.1088/1757-899X/1062/1/012025

Hone, K.S., El Said, G.R.: Exploring the factors affecting MOOC retention: a survey study. Comput. Educ. **98**, 157–168. Elsevier Ltd (2016). https://doi.org/10.1016/j.compedu.2016.03.016

Idrizi, E., Filiposka, S., and Trajkovik, V. 2018. "VARK Learning Styles and Online Education: Case Study," Researchgate.Net (September). (https://www.researchgate.net/profile/Vladimir-Trajkovik/publication/327869001_VARK_Learning_Styles_and_Online_Education_Case_Study/links/5baa663a299bf13e604c7ea7/VARK-Learning-Styles-and-Online-Education-Case-Study.pdf)

Instruction. UNESCO, International Bureau of Education. Accessed 20 Apr 2023. https://www.ibe.unesco.org/en/glossary-curriculum-terminology/i/instruction#:~:text=The%20creation%20and%20implementation%20of,%2C%20attitudes%2C%20appreciations%20and%20values

Jordan, K.: Massive open online course completion rates revisited: assessment, length and attrition. Int. Rev. Res. Open Distrib. Learn. **16**(3), 341 (2015). https://www.erudit.org/en/journals/irrodl/1900-v1-n1-irrodl04980/1065985ar/abstract/

Krashen, S.D., Jones, C.M., Zelinski, S.J., Usprich, C.: How important is instruction? ELT J. **XXXII**(4), 257–261 (1978). https://doi.org/10.1093/ELT/XXXII.4.257

Langley, D.J., Bijmolt, T.H., Ortt, J.R., Pals, N.: Determinants of social contagion during new product adoption. J. Prod. Innov. Manag. **29**(4), 623–638 (2012). https://doi.org/10.1111/j.1540-5885.2012.00929.x

Liu, M., et al.: Understanding MOOCs as an emerging online learning tool: perspectives from the students. Am. J. Dist. Educ. **28**(3), 147–159 (2014). https://doi.org/10.1080/08923647.2014.926145

Liu, S., Liang, T., Shao, S., Kong, J.: Evaluating localized MOOCs: the role of culture on interface design and user experience. IEEE Access **8**, 107927–107940 (2020). https://doi.org/10.1109/ACCESS.2020.2986036

Liyanagunawardena, T.R., Williams, S., Adams, A.A.: The impact and reach of MOOCs: a developing countries' perspective. eLearn. Papers, 38-46 (2014). http://elearningeuropa.info/en/download/file/fid/26956

Ma, L., Lee, C.S.: Investigating the adoption of MOOC s: a technology–user–environment perspective. J. Comput. Assist. Learn. **35**(1), 89–98 (2019). https://doi.org/10.1111/jcal.12314

Ma, L., Lee, C.S.: Drivers and barriers to MOOC adoption: perspectives from adopters and non-adopters. Online Inf. Rev, **44**(3), 671–684 (2020). https://doi.org/10.1108/OIR-06-2019-0203

Moore, M.G.: Three types of interaction. Am. J. Dist. Educ. **3**(2), 1–7 (1989). https://doi.org/10.1080/08923648909526659

Pasha, A., Abidi, S.H., Ali, S.: Challenges of offering a MOOC from an LMIC. Int. Rev. Res. Open Distrib. Learn. **17**(6), 221–228 (2016). https://doi.org/10.19173/irrodl.v17i6.2696

Romero-Rodríguez, L.M., Ramírez-Montoya, M.S., Aguaded, I.: Determining factors in MOOCs completion rates: application test in energy sustainability courses. Sustainability **12**(7), 2893 (2020). https://doi.org/10.3390/su12072893

Rosé, C.P., Ferschke, O.: Technology support for discussion based learning: from computer supported collaborative learning to the future of massive open online courses. Int. J. Artif. Intell. Educ. **26**, 660–678 (2016). https://doi.org/10.1007/s40593-016-0107-y

Sallam, M.H.: A review of MOOCs in the Arab world. Creat. Educ. **8**(04), 564 (2017). https://doi.org/10.4236/ce.2017.84044

Sanchez-Gordon, S., Luján-Mora, S.: Research challenges in accessible MOOCs: a systematic literature review 2008–2016. Univ. Access Inf. Soc. **17**(4), 775–789 (2018). https://doi.org/10.1007/s10209-017-0531-2

Shah, J., Khanna, M., Patel, H.: Technology adoption and continuance of MOOCs: A systematic literature review synthesis, and future research agenda. J. Content Commun. Commun. **14**(7), 106–132 (2021). https://doi.org/10.31620/jccc.12.21/10

Steyn, R.: How many items are too many? an analysis of respondent disengagement when completing questionnaires. Afr. J. Hosp. Tour. Leisure **6**(2), 1–11 (2017)

Svendsen, G.B., Johnsen, J.A.K., Almås-Sørensen, L., Vittersø, J.: Personality and technology acceptance: the influence of personality factors on the core constructs of the technology acceptance model. Behav. Inf. Technol. **32**(4), 323–334 (2013). https://doi.org/10.1080/0144929X.2011.553740

Taherdoost, H.: A review of technology acceptance and adoption models and theories. Procedia Manuf. **22**, 960–967 (2018). https://doi.org/10.1016/j.promfg.2018.03.137

Tang, H.: Person-centered analysis of self-regulated learner profiles in MOOCs: a cultural perspective. Educ. Technol. Res. Dev. **69**(2), 1247–1269 (2021). https://doi.org/10.1007/s11423-021-09939-w

Tornatzky, L.G., Fleischer, M., Chakrabarti, A.K.: The Processes of Technological Innovation. Lexington Books (1990)

Tseng, T.H., Lin, S., Wang, Y.S., Liu, H.X.: Investigating teachers' adoption of MOOCs: the perspective of UTAUT2. Interact. Learn. Environ. **30**(4), 635–650 (2022). https://doi.org/10.1080/10494820.2019.1674888

Urval, R.P., Kamath, A., Ullal, S., Shenoy, A.K., Shenoy, N., Udupa, L.A.: Assessment of learning styles of undergraduate medical students using the VARK Questionnaire and the influence of sex and academic performance. Adv. Physiol. Educ. **38**(3), 216–220 (2014). https://doi.org/10.1152/advan.00024.2014

Venkatesh, V., Thong, J., Quarterly, X.X.-M.: Consumer acceptance and use of information technology: extending the unified theory of acceptance and use of technology. JSTOR (2012). https://www.jstor.org/stable/41410412

Wang, Y., Baker, R.: Grit and intention: why do learners complete MOOCs? Int. Rev. Res. Open Distrib. Learn. **19**(3), 20–42 (2018). https://doi.org/10.19173/irrodl.v19i3.3393

Wu, B., Chen, X.: Continuance intention to use MOOCs: integrating the technology acceptance model (TAM) and task technology fit (TTF) model. Comput. Human Behav. **67**, 221–232 (2017). https://doi.org/10.1016/j.chb.2016.10.028

Zeithaml, V.A.: Consumer perceptions of price, quality, and value: a means-end model and synthesis of evidence. J. Mark. **52**(3), 2–22 (1988). http://www.jstor.org/stable/1251446

AI4T: A Teacher's Dashboard for Visual Rendering of Students' Assignments in Massive Open Online Courses

Filippo Sciarrone[1(✉)], Francesco Paolo Sferratore[2], and Marco Temperini[3]

[1] Faculty of Technological and Innovation Sciences, Universitas Mercatorum, Piazza Mattei, 10, Rome, Italy
filippo.sciarrone@unimercatorum.it
[2] Pegaso University, Piazza Trieste e Trento, 48, Naples, Italy
[3] DIAG-Department of Computer, Control and Management Engineering, Sapienza, University of Rome, Rome, Italy
marte@diag.uniroma1.it

Abstract. The COVID-19 pandemic has changed the way we do education in recent years. In fact, thanks in part to the progress of the Internet, there has been an exponential growth in courses delivered in distance mode. Among these, Massive Open Online Courses are undoubtedly those courses where the growth in enrolments has been strongest: in fact, even in universities there are distance courses with thousands of enrolments. In this scenario, it is really difficult, if not impossible, for a teacher to monitor the learning process of her/his class, unless he or she is equipped with one or more tools enabling him or her to follow the students, in their learning process, in a more analytical manner. In this paper we propose a web tool, the AI4T system, a dashboard usable as a web application, which allows the teacher, once an assignment has been assigned to her/his students, to monitor their outcomes through a representation in a two-dimensional space. We present an initial experiment with encouraging results.

Keywords: Learning Analytics · Dashboard · MOOCs · Deep Learning

1 Introduction

History of Massive Open Online Courses (MOOCs) starts in 2007, when a networked course, opened to anybody able to connect to it via internet, was established at the Utah University, in the United States [10]. Then the name MOOC started to be popular, and be associated to the idea of an educational experience taken online and free to join in, in 2008, with the availability of the course *Introduction to Artificial Intelligence* from Stanford University (USA again) [11]. Since then, the idea of MOOC has developed widely, also having a further impulse in the very last years, from the sudden, and often haphazard, emergency

H. Xie et al. (Eds.): ICWL 2023, LNCS 14409, pp. 13–27, 2023.
https://doi.org/10.1007/978-981-99-8385-8_2

passage from face to face to online education, originated by the COVID-19 experience we had to endure. However, the trend to have a large growth in time was already established in the previous years: from 2016 to 2019, the number of students enrolled in MOOCs, the number of universities hosting at least a MOOC, and the number of courses given as MOOCs, grew, respectively from 58 millions, 700, and 6.9 thousands, to 120 millions, more than 900, and 13.5 thousands. In 2020 the figures were, respectively 180 millions, 950, and 16.3. In 2021 they were 220 millions, 950, and 19.4 thousands[1] In time, the original openness and *free-to-join* attitude of MOOCs have quite vanished. However, a vast amount of research was conducted about MOOCs [4]. These research activities revealed MOOCs potentialities for effective support to distance learning, and showed how MOOCs can complement the face to face educational activities, also profiting of techniques and methods of the Technology Enhanced Learning (TEL) research area and in particular by Learning Analytics (LA) [14,15]. Assessment is one of the challenges in a MOOC's management. Due to the large amount of enrolled students, the teacher is confronted with the task of evaluating big numbers of students assignments. Often the solution is in increasing the number of teachers or tutors, or to decrease the complexity and significance of the homework, by a large use of multiple choice questionnaires. However, there are innumerable skills that cannot be so easily assessed, whereas they require an analysis of the individual student's product, which is impractical when big numbers of students are implied, and several assignments have to be assessed in a semester. This difficulty hinders significantly the teacher's capability to keep under control the learning processes of all their learners [8,16]. At least for homework mainly based on the use of language (essays), Deep Learning (DL) and in particular Large Language Models (LLMs), can be used to support the teacher in the assessment activity. In this paper we present the AI4T system aiming to provide the teacher with a dashboard consisting in a visual representation of a MOOC's essays, as an instrument to analyze the class products, and for the selection of individual essays to be directly analyzed. We use the *Text Embedding* Deep Learning (DL) technique, to represent each essay through a numerical dense vector, as opposed to the classic representations based on the *Bag of Words* approach [12]. In our case each embedding is composed by 700 real numbers. We use a LLM, the *doc2vec* DL model [7], to compute the essays embeddings, plus, if available, a teacher's version of the essay, to be used as the baseline for comparison and classification. Subsequently, on each embedding we apply a dimensionality reduction, through the *Principal Component Analysis* (PCA) technique [5], to represent it as a point in a 2-D space. Since the embedding brings the semantic characterization of the essay, essays that are approximately in the same area of the 2-D space are likely to be similar in their content [7]. While we do not go into the automated analysis of correctness of the essay's content, we allow the teacher to: 1) see how the essays (points) are spread in the space, and 2) *open* those documents that (s)he considers worth to analyse. For instance, the teacher might be interested in checking on those essays that appear to be far away with

[1] According to https://www.classcentral.com/report/.

respect to her/his own, to see whether that distance was due to inconsistent contents, or to uncommon observations. To check the usefulness of the system, we submitted a questionnaire to a sample of 18 teachers, selected from both high school and university setting and, by using ChatGPT[2], simulated a MOOC class, with 200 students and the related essays, imagining to deal with a task-session proposed to the students, involving the submission of an essay each. Then we proposed the 2-D visualisation of the essays, as produced by the system, to the sample, by the questionnaire deemed to investigate about the usefulness that the responding teachers were seeing into the eventual usage of the system. In terms of a Research Question (RQ), we intended to check: *how useful the dashboard could be for the teacher in terms of comprehension of the learning process of the class, as it appears from the 2-D rendering.* In the following sections we discuss related literature (Sect. 2), present the general architecture of the system (Sects. 3 and 4), present and discuss the results of the questionnaire (Sect. 5), and draw some concluding remarks in (Sect. 6), also to point out some limits of this research and some future work.

2 Related Work

LA is a multi-disciplinary field involving models based on machine learning, information retrieval, statistics, and visualization. Several models of LA have been proposed in the literature in order to holistically support the process of teacher inquiry [13,17,18]. In this context, in last years, dashboards have become a popular means to present critical information at a glance. Generally speaking, a dashboard is a visual display of the most important information needed to achieve one or more objectives; consolidated and arranged on a single screen so that the information can be monitored at a glance. LA is best realized with dashboards including meaningful information about learning tasks and progress of learning towards specific goals: students who received dashboard analysis obtained a higher final score than those who did not [3]. The dashboard presented in [9] is a dashboard for analyzing python programs. The authors address, by the same DL technique, the writing styles differences of Python programs proposed by a students belonging to a MOOC. So their work is in the framework of *Source Code Embedding*, that is the DL attempt for supporting automatic *Source Code Analysis* [1,6]. The research question was to relate the distance between two python programs in the 2-D space to their dissimilarities in the writing style. Our proposal is more general. Firstly because here we propose a complete web application, shown in the next two sections, and secondly because we apply the DL approach to a more general goal, not necessarily related to science or IT disciplines, but potentially usable in any field.

In [2] a LA dashboards to support adviser-student dialogue is proposed. They present *LISSA* (Learning dashboard for Insights and Support during Study Advice), a learning analytics dashboard designed, developed, and evaluated in collaboration with study advisers with the main goal to facilitate communication

[2] https://chat.openai.com/.

between study advisers and students by visualizing grade data that is commonly available in any institution. More specifically, the dashboard attempts to support the dialogue between adviser and student through an overview of study progress, peer comparison, and by triggering insights based on facts as a starting point for discussion and argumentation. This system does not use DL to represent students' data but it's rather a graphic representation of students' grades during their learning process. Our system aims at helping teachers analyse homework (see Sect. 5).

In [14], a dashboard is presented, called *K-OpenAnswer*, which helps teachers to simulate the dynamic of a MOOC where peer assessment is used. The system uses a machine learning technique, based on a modified version of the K-NN algorithm, and provides teachers with a statistical and graphic environment by which they can monitor the evolving dynamic of a simulated MOOC. An experimental evaluation is presented that highlights the advantages of using this dashboard as a valid tool for the study of real MOOCs. This dashboard works on simulated data, acting as a what-if analysis for the teacher. Here each simulated student's model is represented by a point in a 3-D space. On the other hand, in our system, each point represents a student's real homework and allows for an in-depth analysis of the MOOCs learning process. The study by Susnjak et al. [19] investigates current approaches to LA dashboarding while highlighting challenges faced by education providers in their operationalization. The authors identified 17 LA dashboards, assessing their common characteristics as well as strengths and weaknesses. They take into account 3 *capabilities*, to analyze them: descriptive analytics, predictive analytics and prescriptive analytics. They propose a dashboard for students as well, to monitor grades, engagement and so on. Differently our dashboard, at least at this stage, is proposed mainly for teachers, giving them a tool for investigating the learning process associated to a particular topic.

3 The Architecture of the AI4T System

In this section the functional architecture of the system is shown. Figure 1 depicts the general functional blocks:

- The Login module;
- The Registration module;
- The Active Sessions Visualization module;
- The Specific Sessions Visualization module;
- The Creation Session module.

The Login and Registration modules provide some basic authentication functionalities, such as the RegEx validation, the SHA-256 cryptography and the e-mail confirmation. In the registration page one can register either as a student or as a teacher, each role having its own workflow and interaction with the application. The Active Sessions visualization consists of two main modules, depending on the role of the user. For the teacher role, the system shows a list of all the

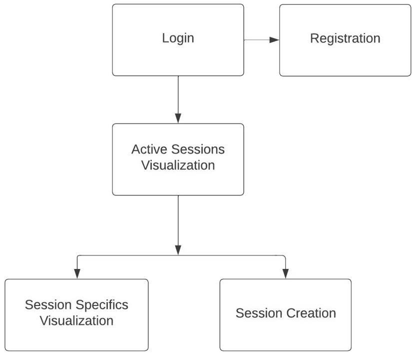


Fig. 1. The general functional blocks composing the AI4T system.

sessions, where each row is the link to the Session Data visualization page. The teacher can launch a new session as well. For the student role, the environment appears different: a *search* bar is shown, where the student can insert the code of an open session to upload her/his essay. The Specific Visualization Session provides the graphical representation of the essays, uploaded by the students and having the same session identification code. This representation is based on *doc2vec* DL model, followed by a PCA-driven dimensional reduction. This module allows teachers to examine each essay by means a suitable file content reader. The content of each essay can be visualized by clicking the correspondent point in the 2-D space. Moreover, the system allows the teacher to upload her/his own version of the requested essay, to be used as the baseline. Such baseline appears in the graphic, as a red dot, while students' essays are represented by blue points. Finally, the Creation Session module consists of a simple web form showing (i) an input field for an auto-generated and not-editable session code; (ii) an input field for the name of the session; (iii) an input field to upload the baseline essay.

4 The AI4T System at Work

As described in the previous section, the system presents two main workflows: one for teachers and one for students. In fact, while the Session Manager creates new task-sessions and allows teachers to analyze data, the students can use the system to upload their essays, connecting them to a particular session. The workflows are described in the following subsection. Common to both is the authentication service, that allows to initiate a login session, or to access the registration form.

The student's workflow is represented in Fig. 2. It can be summarized as follows, after logging in:

1. The student enters the session code, provided by the teacher, into the Search Bar in the Active Session module. The session code is then transmitted to the Controller module;

2. The Controller module retrieves the session data from the local database: if the session exists, the controller sends back the Specific Session html page;
3. The student uploads her/his essay through a suitable file uploader;
4. The Controller stores the essay into Session folder.

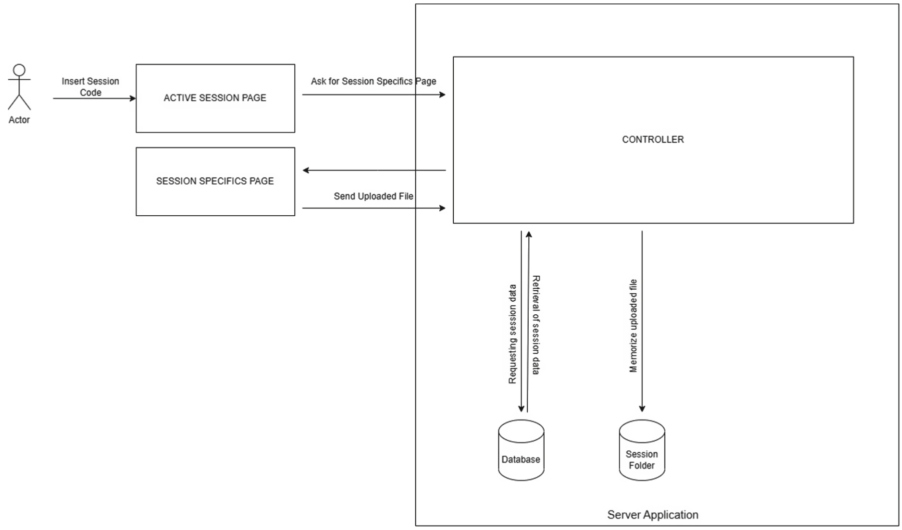


Fig. 2. The student's workflow.

In Fig. 3, the teacher's workflow is shown. After logging in, the workflow is based on the following steps:

1. The teacher clicks the "SHOW DIAGRAM" button on the Specific Session page. After that, the session ID is transmitted to the Controller;
2. The Controller receives the request and launches the PROCESS_DATA function;
3. The PROCESS_DATA function:
   – Retrieves all the essays related to the selected session from the local MySql database, the database containing users' and sessions' data;
   – Formats the essays in such a way to be processed;
   – Retrieves the hyper-parameters needed to the doc2vec DL model. In this case we used a model based on an embedding composed by 700 components;
   – Processes the essays using the doc2vec DL model;
   – Launches the PCA module for the dimensionality reduction process: each embedding is transformed into a 2-dimensional array, which represents a single point into a 2-D space;
   – Stores all the points into a local json files.

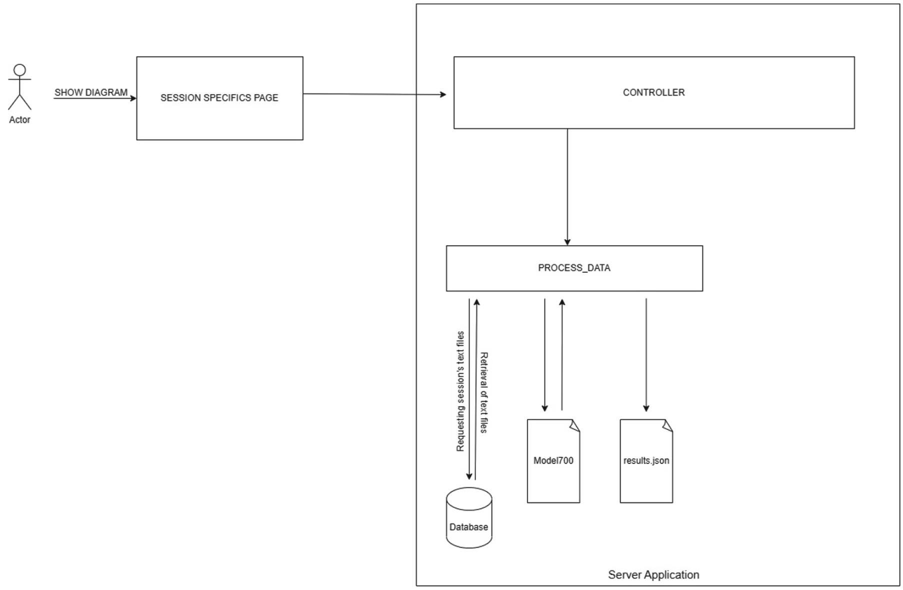


Fig. 3. The teacher's workflow.

4. The CONTROLLER module manages the json file to provide a graphic representation of all the essays within the Session Manager environment.

Figures 4 and 5 show the straightforward interaction system-teacher, at the time of task-session management and creation.

5 Investigation on the Usefulness of the System

To answer the RQ we proceeded in two steps, we: 1) simulated a MOOC's task session, composed by the essays submitted by 200 students, plus the one uploaded by the teacher; 2) submitted a questionnaire to the sample of 18 teachers, as mentioned in Sect. 1, mainly to verify how they would consider the system visualisation useful, and, in particular, how they agreed on the usefulness of certain didactic applications we suggested for such a kind of visualisation. We also were asking for free-text comments, with observations and possible further uses of the system, in each one of the questions, and in a final open-answer.

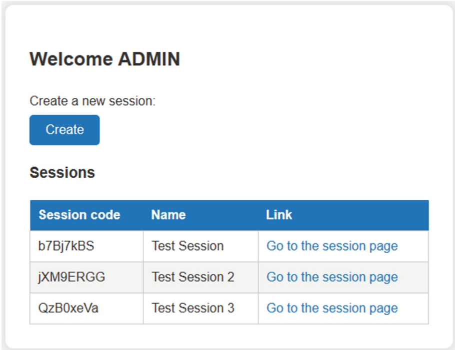


Fig. 4. The Teacher's workflow: a list of available task-sessions is shown. Each one calls the students to answer, by producing their own essays.

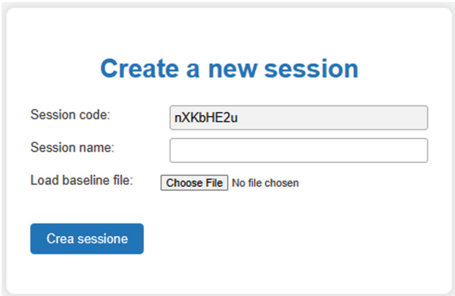


Fig. 5. The Teacher's workflow: creation of a new task session, by assigning a name and providing, optionally, the exemplar essay coming from the teacher.

5.1 The Simulation Procedure

The essay to be produced by the students was the following: *Write an essay to illustrate the figure of Napoleon in a maximum of* 250 *words*. To produce the sample of 200 essays, we used ChatGPT[3], as the essay generator, through the following procedure:

1. ChatGPT was initially prompted with the following statement: *Write an essay to illustrate the figure of Napoleon in 250 words;*
2. Nine semantically-close texts were reproduced. ChatGPT was prompted nine times with the following statement: *Now generate another one using different words and different order of explanation of the concepts;*
3. Other 190 essays were produced using the following statement: *Now generate ten essays with different concepts and ideas compared to the previous one. Use a maximum of 250 words.*

[3] chat.openai.com.

We used this procedure just to simulate a MOOC, where this approach could be more useful with respect to a normal class with few students. However, this approach does not affect the system evaluation, because here we investigate the first step, i.e., the usefulness of the system from the teachers' point of view and not other functionalities like usability, and happiness from students' point of view. So, the simulation here acts as a *What-If analysis*. Hence, after processing the essays, we obtained their 2-D representation, as shown in Fig. 6. Figure 7 shows the same distribution, where the teacher has clicked on his/her own essay. In fact, by clicking on each point, the corresponding essay is visualized.

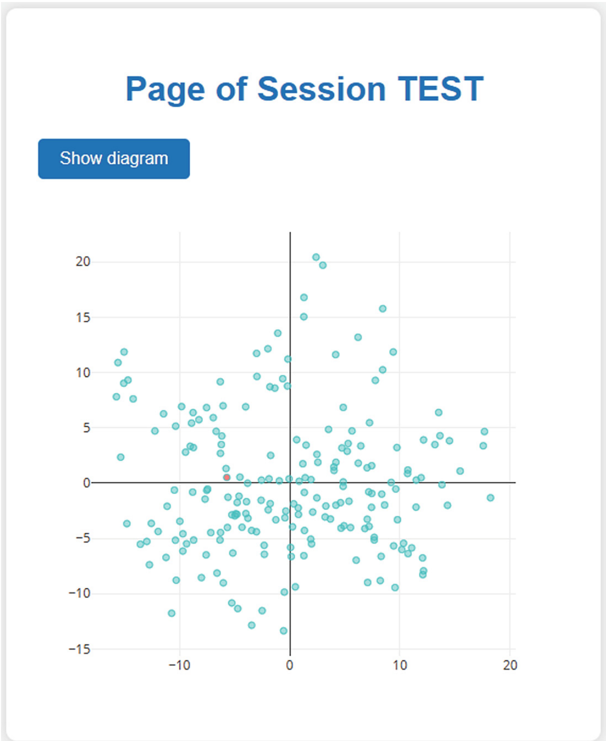


Fig. 6. The essays 2-D representation. The quadrants of the graph have no information content: they have been highlighted just to help visually divide the areas of the graph.

5.2 The Usefulness Questionnaire

In order to evaluate the system from the teachers' point of view, a questionnaire, composed by six questions, was submitted to the sample. The first five were based on a 5-points Likert scale, ranging from *Strongly disagree*, to *Strongly agree*, while

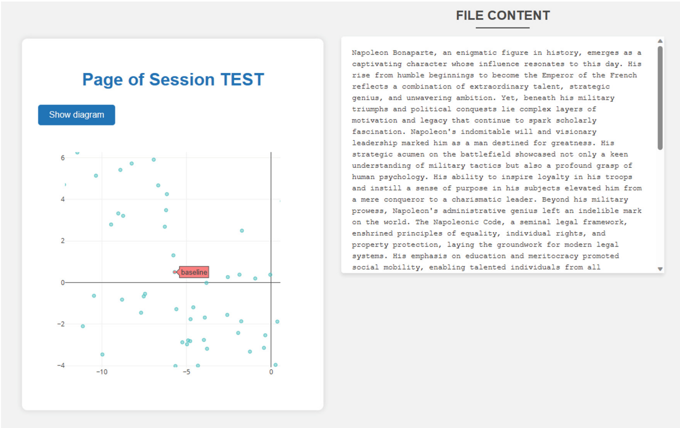


Fig. 7. The teacher can analyze each essay by clicking on its point in the 2-D space.

the last one was a free-text question for comments and suggestions. Table 1 shows the questions. All the questions were about Fig. 6 and Fig. 7, shown to the sample at the beginning of the questionnaire.

Table 1. The questions. They were introduced by an explanation of the visualization system, based on the pictures reported in Fig. 6 and Fig. 7.

Q1	The 2-D representation, shown above, is useful, as it allows the teacher to have a possibly interesting visualization of a student's learning process
Q2	The 2-D representation, shown above, is useful, as it allows the teacher to have a possibly interesting visualization of the learning process of the whole class of students
Q3	A point distant from the teacher's point might of course just be wrong (not fulfilling the answer requirements), however, it might also represent an essay containing creative and novel concepts (though different from those in the exemplar essay)
Q4	If the 2-D representation showed a group of points (essays) being 1) close to each other, and 2) far from the teacher's one, it could be useful to examine at least one of them
Q5	Two essays that are represented by two quite close points could be evaluated by the same grade
Q6	Do you see other possible useful applications of the 2-D representation? Please add your comments and suggestions here

5.3 Results

The results of the questionnaire are summarized in Fig. 8, where, for each question the weighted average is shown, having managed the Likert scale as a continuous variable.

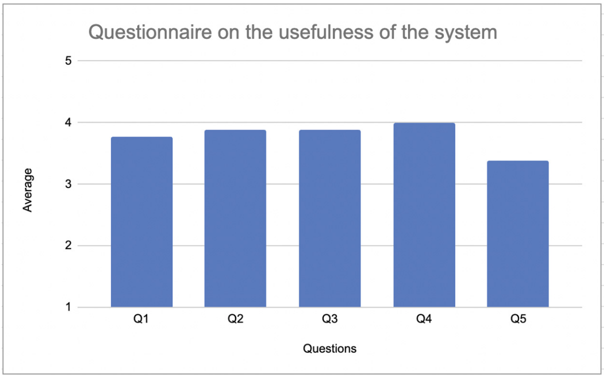


Fig. 8. The questionnaire's results.

5.4 Discussion

The answers to the questionnaire give a valuable feedback about the uses and enhancements of the AI4T system. In addition to the average values, shown in Fig. 8, we also provide a different view of the frequencies, grouping together the values 4 (Agree) and 5 (Strong agree) as confirming the teachers' interest in the use of the system, and the values 1 (Strongly disagree) and 2 (Disagree) as a negative feedback. We do not take into account the neutral value, excluding it from the assessments.

Question Q1. The first question was related to the possibility, for the teacher, to appreciate an indication about the state of the individual student learning process. Of course, in this case, such appreciation would come not only from the visualisation of a whole session, but also from the comparison of that individual student position in the graphics drawn for the set of task-sessions in which (s)he participated. In this regard the answers were quite positive, having twelve confirming and one not confirming: one occurrence of 2, five of 3, nine of 4, and three of 5. So, the idea of following the outcome of an individual student, possibly along different sessions, seems to be attractive. On this question we received some comments. Among others, such as centering the teacher's exemplar essay in the graphic, or partitioning the space in concentric areas around that red dot, one was particularly evocative for us: *perhaps more than the red dot with the teacher's essay, the algorithm could automatically identify several red dots (centroids)*: in this case the grouping of students/essays in clusters seems to be considered with interest, as a way to check only some essays, as representative of several more.

Question Q2. This question generalized to the whole MOOC the rationale behind the previous one. Would a glance at the spatial distribution of students/essays/points be evocative of a *learning state of the class*? In this regard the answers are encouraging, with five occurrences of 3, ten of 4, and three of

5. Hence, the MOOC's 2-D representation provided by the system, turned out to be an interesting tool for monitoring the learning process. We received two relevant comments. The first suggested to check the system's functionality by comparing its behavior to different typologies of MOOCs, differentiated by average prowess on the subject matter. We consider this comment as suggesting a different basis for the visualisation. In particular, also in view of some of our reflections coming during the development of the system, we can connect this comment to the idea of managing the presence or absence of important concepts in the students' essays, as an alternative way to distribute the points on the plane. In other words, a different meaning of distance, with respect to the one is used here. The second comment was about the problems introduced by the dimensional reduction which the graphic is based on. As a matter of fact, we had anticipated the above caveat since the start of this work, and we think that it can be eased only by further work, such as determining the influence of the features compressed by the PCA technique on the conservation the essays' semantics.

Question Q3. This question was imagined as a way to check the limits of our approach: if a point is distant from the teacher's one, we cannot just condemn it as *wrong*. We think that the responding teachers were quite in agreement with the need to check those points (essays) anyway. Q3 had 13 confirming values and 1 negative feedback. The results are: one occurrence of Likert value 2, four occurrences of 3, nine of 4, and four of 5. Comments were quite valuable. In one comment, out of four, the fact that only the teacher could distinguish between *creative/critical thinking*, and *wrong* was considered as a limit. About this we could add that our aim in this work is to give rise to a system that could orient the teacher in the *space of the essays* rather than provide an evaluation means. However, in our interpretation, the interest about a tool that could possibly extract more meaningful features from an essay and help the teacher cut further the analysis of a big number of essay, is a compelling topic for future work. Especially in view of the other suggestions about possible automated verification of presence/absence of concepts/topics in each essay, which is basically the content of the other comments to this question.

Question Q4. The last two questions, Q4 and Q5, were devised to delve into the possible significance of the system's graphics for assessment aims. Although we are proposing the current development of the system as an *orientation tool*, rather as an *evaluation tool*, we considered that the second interpretation would be a natural consequence in the eyes of a teacher compared with hundreds of essays (or more), so we thought the topic should be met in some way right away. In particular Q4 suggests the idea that looking at one, or some, of the essays could help classifying the other essays that are close by. There is an agreement on that, by the sample, if less high as for the previous questions: 12 confirming values and 2 negative feedback, with two occurrences of 2, four of 3, four of 4, and eight of 5, In one case, the comment mentioned the possibility that a cluster of essays might reveal a shared misconception, and help deciding whether to provide remedial activities. We interpreted this comment as an additional suggestion about the management of concepts extracted from the essays, which is a different way to measure essays features and place them in the 2D space, different from the one we are currently using.

Question Q5. The last question was more explicit, in relation to the possibility of turning the system into an evaluation tool. Surprising enough, we had no hugely negative feedback, although this is the most controversial set of answers in the questionnaire: 8 confirming values and 4 negative feedback, with four occurrences of Likert value 2, six of value 3, five of value 4, and three of value 5, The responding teachers were rightly suspicious about this direct usage of the system: one, out of three comments, confirmed that anyway the teacher should check, before giving green light to grading.

6 Conclusions, Limits of the Research, and Future Work

In this article, we presented the AI4T system, available as a web application, with the aim of giving teachers a dashboard to hold and monitor task sessions based on the production of individual students' essays. The tool is deemed to be a useful asset for the analysis of essays in the framework of a MOOC, where the number of elements could make the guidance offered by the system valuable. We showed the system to a sample of teachers, at high school and higher education levels, and obtained a set of answers. We analysed the answers and could draw here some conclusions, related to 1) the usefulness of the system, perceived by the responding teachers, and 2) the limitations of our current approach. From the answers and from the comments in the questionnaire we concluded that the responding teachers were in quite strong agreement with the idea that our system can be useful as a guidance to the analysis of the essays. In this respect we think we got a positive result, in view of the actual use of our system, after further refinements. We also concluded that there is a natural need for the support to automated grading of the essays, that our system 1) was not meant to give so far, and 2) is not apt to give in its current implementation. On the other hand, the investigation was, in turn, useful to confirm/unveil some limits of the current system and some occasions for further enhancements.

6.1 Limits of the Research

The comments of the sample allowed us to reflect on two main limits:

- One is about the concept of distance, as connected to the semantics of the essays. The method we use to visualise the essays in a 2D space should be put on trial with further evaluation, in order to convincingly allow to consider two close points as representative of so semantically close essays as to let them share the grade (or part of it).

- Related to the previous item, the system shows the essays/points in a space divided by axes, whose meaning cannot be made clear to the user;
- The system has no alternative method of visualisation that could help appreciating semantic similarity of the essays, so it has to rely on the only one method we have demonstrated in this paper.

6.2 Future Work

In terms of future work we may point out one of the developments that we think could ease the limits earlier exposed, and, to some extent, help to make the visualisation more selective and useful, in view of some of the free comments provided by the responding teachers. So an immediate aim for future work is about the alternative visualisation mentioned in the third limit of this research. In particular, the management of concepts treated in the essays would help on this point. Basically, the extraction of concepts/topics from the essays (students' and teacher's) could allow analysing the space of points/essays from a different viewpoint: inclusion or lacking of important concepts, with a consequent different idea of discrete distance among the points. This could help the teacher to see, in the graphics, a second distributions of the points, different than the one currently offered: for instance, the distant points would be essays where the expected concepts have been either missed, or substituted by different ones.

Acknowledgements. This work was supported by the funding project AI4DIX-16-FIN/RIC from Universitas Mercatorum.

References

1. Alon, U., Zilberstein, M., Levy, O., Yahav, E.: Code2Vec: learning distributed representations of code 3(POPL) (2019)
2. Charleer, S., Moere, A.V., Klerkx, J., Verbert, K., De Laet, T.: Learning analytics dashboards to support adviser-student dialogue. IEEE Trans. Learn. Technol. **11**(3), 389–399 (2018)
3. Ifenthaler, D., Yau, J.K.: Utilising learning analytics to support study success in higher education: a systematic review. Education Tech. Research Dev. **68**, 1961–1990 (2020)
4. Irwanto, I., Wahyudiati, D., Saputro, A., Lukman, I.: Massive open online courses (MOOCs) in higher education: a bibliometric analysis (2012–2022). IJIET: Int. J. Inf. Educ. Technol. **13**(2), 223–231 (2023)
5. Jackson, J.E.: A User's Guide to Principal Components. Wiley-Interscience Paperback Series. Wiley, New York (2005)
6. Kanade, A., Maniatis, P., Balakrishnan, G., Shi, K.: Learning and evaluating contextual embedding of source code. In: International Conference on Machine Learning (ICML), Vienna, Austria (2020)
7. Le, Q., Mikolov, T.: Distributed representations of sentences and documents. In: Proceedings of the 31st International Conference on International Conference on Machine Learning, ICML 2014, vol. 32, pp. II-1188-II-1196. JMLR.org (2014)

8. Limongelli, C., Sciarrone, F., Temperini, M., Vaste, G.: Lecomps5: a web-based learning system for course personalization and adaptation. In: Proceedings of IADIS International Conference e-Learning, pp. 325–332 (2008)
9. Mastrostefano, S., Sciarrone, F.: Monitoring programming styles in massive open online courses using source embedding. In: 2022 26th International Conference Information Visualisation (IV), pp. 245–250 (2022)
10. Mota, R., Scott, D.: Education for Innovation and Independent Learning. Elsevier, New York (2014)
11. Mutawa, A.M.: Perspective chapter: MOOCs at higher education: current state and future trends. In: Goundar, D.S. (ed.) Massive Open Online Courses - Current Practice and Future Trends, Chapter 3. IntechOpen, Rijeka (2023)
12. Piech, C., Huang, J., Nguyen, A., Phulsuksombati, M., Sahami, M., Guibas, L.: Learning program embeddings to propagate feedback on student code. In: Bach, F., Blei, D. (eds.) Proceedings of the 32nd International Conference on Machine Learning, Proceedings of Machine Learning Research, vol. 37, pp. 1093–1102. PMLR, Lille, France (2015)
13. Sciarrone, F., Temperini, M.: Learning analytics models: a brief review. In: Proceedings of the 23rd International Conference on Information Visualisation (IV-2019), vol. 2019-July, pp. 287–291 (2019)
14. Sciarrone, F., Temperini, M.: K-OpenAnswer: a simulation environment to analyze the dynamics of massive open online courses in smart cities. Soft. Comput. **24**(15), 11121–11134 (2020)
15. Sciarrone, F., Temperini, M.: A web-based system to support teaching analytics in a MOOC's simulation environment. In: Proceedings of the International Conference on Information Visualisation, vol. 2020-September, pp. 491–495 (2020)
16. Sciarrone, F., Temperini, M.: Monitoring massive open online courses (MOOC) during the Covid-19 pandemic. In: Visvizi, A., Troisi, O., Saeedi, K. (eds.) RIIFORUM 2021. SPC, pp. 143–153. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-84311-3_14
17. Sergis, S., Sampson, D.G.: Teaching and learning analytics to support teacher inquiry: a systematic literature review. In: Peña-Ayala, A. (ed.) Learning Analytics: Fundaments, Applications, and Trends. SSDC, vol. 94, pp. 25–63. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-52977-6_2
18. Sterbini, A., Temperini, M.: Supporting assessment of open answers in a didactic setting. In: Proceedings of the 12th IEEE International Conference on Advanced Learning Technologies, ICALT 2012, pp. 678–679 (2012)
19. Susnjak, T., Ramaswami, G., Mathrani, A.: Learning analytics dashboard: a tool for providing actionable insights to learners. Int. J. Educ. Technol. High. Educ. **19**(12), 1–23 (2022)

Corpus-Based Translation Pedagogy: A Preliminary Case Study

Mengqi Li(✉) and Dongxia Pan

School of Foreign Studies, Anhui Xinhua University, Hefei, AnHui 230088, China
mengqili2022@gmail.com, pandongxia@axhu.edu.cn

Abstract. This research is grounded in the practical experiences of teaching translation at a specific local university and aims to introduce an innovative corpus-based pedagogical framework tailored for translation courses. Through a comparative study involving undergraduate students' training experiments, the research seeks to provide valuable insights into the effective integration of corpus technology in translation education, meticulously examining aspects such as course design, corpus resources, teaching content, and implementation strategies. The results of the teaching practice reveal that the newly proposed pedagogical model significantly enhances the expression of translations, elevates students' English proficiency as translators, diversifies learning methods, and fosters increased interest and self-learning capabilities among students. Building upon these findings, future translation education should prioritize the enrichment of corpus teaching resources and dynamic tracking of students' daily corpus usage. Translation instructors must continually refine their corpus usage skills and enhance their digital integrated capabilities to effectively adapt to the evolving needs of translation course reforms in the current context. This research contributes to the ongoing discourse on the integration of corpus technology and Data-Driven Learning (DDL) in language teaching, offering insights into how educators can harness the potential of corpus tools and DDL pedagogy in their classrooms.

Keywords: Corpus-based Translation Pedagogy · Corpus Technology Integration · Data-Driven Learning (DDL)

1 Introduction

In the ever-evolving landscape of language education, the fusion of technology and pedagogy has redefined how we approach teaching and learning. One of the forefronts of this transformation is Corpus-Based Translation Pedagogy, which offers a promising avenue to enhance the quality and efficacy of translation education.

Our research is rooted in practical experiences gained from the realm of translation teaching at a specific local university, and it embarks on a journey to reshape the landscape of language education. This exploration unlocks invaluable

H. Xie et al. (Eds.): ICWL 2023, LNCS 14409, pp. 28–38, 2023.
https://doi.org/10.1007/978-981-99-8385-8_3

insights into the seamless integration of corpus technology into the domain of translation education.

In the pages that follow, we will present the findings of a comparative study involving undergraduate students' training experiments. These findings offer valuable insights into the effective integration of corpus technology into translation education. The empirical results of our teaching practice reveal a compelling narrative: the adoption of our newly proposed teaching model significantly enhances translation expression, elevates students' English proficiency as translators, introduces diversity into learning methods, and nurtures heightened interest and self-learning capabilities among students.

The objective of this paper is clear: to craft a corpus-based pedagogical framework that is finely tuned for translation courses. This framework, meticulously designed through a process of inquiry and experimentation, holds the potential to revolutionize translation instruction.

In conclusion, this study significantly contributes to the ongoing discourse surrounding the fusion of corpus technology and Data-Driven Learning (DDL) in language teaching. It serves as a guiding torch for educators, illuminating the path toward acquiring the necessary tools to harness the potential of corpus tools and DDL pedagogy within their classrooms.

2 Literature Review

2.1 Corpus-Based Translation Pedagogy Research

Since 1993, scholars such as Mona Baker, Stewart, Zanettin, Lopez-Rodrigues, and Tercedor-Sanchez have explored the application of monolingual corpora, parallel corpora, and other corpus-based resources in translation research and teaching. Their studies have highlighted the positive impact of corpora on translation pedagogy while also addressing the potential and limitations of integrating corpora into translation instruction. Stewart [12] recognized the role of monolingual corpora in translation teaching. Lopez-Rodrigues and Tercedor-Sanchez [5] developed learner corpora and reference corpora to assist students in mastering translation strategies and improving their evaluation skills. Zanettin [15] emphasized the importance of teachers investing sufficient time in guiding students to query corpus data and analyze the information obtained.

In addition, Shih and Shen [11] pointed out that the use of corpora allows for a comparison between the works of professional translators and student translators, suggesting its usefulness in teaching. Such corpora enable students to observe the differences between themselves and professional translators, rather than merely focusing on their own translation errors.

These studies provide compelling evidence of the application value of corpora in teaching and learning. They highlight the advantages of using corpora as a tool for enhancing students' translation competence, language skills, and language awareness. By incorporating corpora into translation pedagogy, instructors can guide students to analyze language data, observe professional translation practices, and bridge the gap between students' own work and that of experienced translators. Such pedagogical approaches foster critical thinking, promote accuracy, and cultivate a deeper understanding of translation processes and strategies.

While scholars have made significant progress in corpus-based translation pedagogy, there is still room for further exploration and innovation. The integration of monolingual and parallel corpora, along with the incorporation of student translation corpora, holds great potential for creating a dynamic and engaging learning environment that fosters students' translation skills, critical thinking, and self-reflection.

Overall, corpus-based translation pedagogy research has consistently demonstrated the positive impact of corpora on translation learning. It is crucial for educators to harness the potential of corpora in order to enhance the effectiveness of translation instruction and foster the development of competent and reflective translators.

2.2 DDL

The use of language corpora for language teaching and learning, commonly referred to as "data-driven learning" (DDL) as proposed by Johns (1991), has emerged as a well-established and vibrant field of research. According to Boulton & Vyatkina's research [1] in 2021, over the years, nearly 500 empirical studies have contributed to the understanding and application of DDL in diverse educational contexts.

DDL typically encompasses two key approaches: the first involves teachers providing printed concordances of pre-selected corpus data for language learners, while the second empowers learners to directly engage with corpus query software. These approaches serve as valuable tools in aiding the acquisition of various target language constructs, including vocabulary, grammar, collocation, and addressing L2 errors. What makes DDL particularly versatile is its applicability across a wide spectrum of learning and teaching contexts, including English as a Foreign Language (EFL) education.

Recent efforts have aimed to harmonize DDL methodology with established language learning theories. Scholars such as L. Flowerdew [2], Pérez-Paredes [8], and O'Keeffe [7] have highlighted that DDL enhances language learning through a usage-based paradigm. This emphasizes the significance of frequency, salience, and contingency, as articulated by Ellis [3] in 2006. The implementation of DDL involves a blend of constructivist, student-led, and focus-on-form pedagogical activities at the individual level, complemented by teacher- and peer-mediated/scaffolded focus-on-form activities at the sociocultural level. These principles form the core theoretical underpinnings of DDL, contributing to its effectiveness in language education.

The research on DDL has taken various dimensions, shedding light on its impact from multiple perspectives. Studies have investigated classroom corpus use through lenses such as learning gains and reduced L2 errors in pre-/post-test experimental studies. Moreover, research has delved into the perceptions of both learners and teachers regarding DDL (e.g., Yoon & Hirvela [14], 2004), corpus usage patterns (e.g., Pérez-Paredes et al. [9], 2011), and increasingly, the outcomes of teacher education workshops (e.g., Chen et al. [2], 2019; Schaeffer-Lacroix, [10] 2019).

To address teachers' concerns and enhance the integration of DDL into language education, scholars have called for increased alignment between DDL and language learning didactics. In 2019, Meunier [6] advocates for "constructive alignment," emphasizing that curriculum, learning outcomes, teaching methods, and assessment practices should consistently and coherently align with DDL. The Technology, Pedagogy & Content Knowledge (TPACK) model proposed by Koehler and Mishra [4] in 2009 serves as a promising framework for this alignment, as it encompasses pedagogical content knowledge. This knowledge extends beyond understanding how students learn language to integrating DDL within the curriculum and lesson planning, which has been relatively underexplored in existing DDL-focused teacher education studies. Therefore, this study aims to explore how DDL may be seamlessly integrated into Computer-Assisted Language Learning (CALL) lesson planning in the context of translation pedagogy at the university level, contributing to the broader adoption of DDL in mainstream education.

3 Current Issues and Reflections on Translation Courses in Higher Education

Translation courses are compulsory subjects for senior students majoring in foreign languages in universities. They are comprehensive courses that combine theory and practice, aiming to apply students' English abilities effectively. However, through teaching practice and research, it has been observed that the practical English application abilities of many students are weak, particularly in terms of English language awareness and discourse knowledge.

In terms of teaching methods and content, traditional teaching methods with a teacher-centered approach have been followed in translation courses in universities for many years. Teachers often determine textbooks and distribute translation materials based on personal preferences. In the classroom, teachers commonly adopt an immediate exercise and on-the-spot feedback model, lacking language awareness training and appreciation of stylistic features. Translations of words and phrases often lack context, and semantic coherence is lacking. The main basis for translation evaluation by teachers is the answer key from reference books. Some teachers also have unclear understanding and grasp of the translated texts, making it difficult for students to be convinced.

In response to the various issues in translation courses in higher education, in recent years, many scholars have proposed various solutions, one of which is the open corpus-based translation teaching model using computer corpora. In 2007, Luo Xuanmin et al. [13] put forward four advantages of the open corpus-based translation teaching method:

1) openness in *teaching ideas and theories* - students should be exposed to various translation phenomena and become familiar with different texts and translations;
2) openness in *teaching platforms* - corpus construction is a dynamic and improving process, and the use and development of corpora can better serve translation teaching;
3) openness in *data-driven learning* - learners actively use word index software to discover, summarize, and induce relevant patterns, thereby stimulating learning interest and forming an autonomous and open learning mode;
4) openness in *translation text evaluation mechanism* - drawing on machine translation software to test students' and teachers' translations in real-time and summarize common issues.

It is recognized that the fourth point involves complex computer technology and software costs, making it challenging to implement. However, the other three advantages are evident and worth exploring by teachers in their teaching practices.

In conclusion, the current issues in translation courses in higher education in China require attention and appropriate solutions. The adoption of an open corpus-based translation teaching model, as proposed by scholars, can address some of the challenges and offer benefits such as exposing students to diverse translation phenomena, utilizing corpus resources, promoting DDL, and implementing effective evaluation mechanisms. Teachers are encouraged to embrace innovative teaching methods and explore the potential of corpus-based translation pedagogy to enhance students' translation competence and overall learning experience.

4 Methods and Objectives

The study examines the students' perspective on the implementation of corpus in English translation classes. The participants received training, followed by guided and independent practice using a freely available corpus. At the conclusion of the semester, an anonymous questionnaire on their experience with DDL was distributed to the students. The aim of this research is to explore the potential of corpus-based approaches in reforming translation courses for English major students.

4.1 Participants

The participants in this study were third-year undergraduate students majoring in English at Anhui Xinhua College. One class consisted of 35 students and served as the control group, while another class consisted of 36 students and served as the experimental group. A total of 71 participants were involved. From March to June 2023, activities such as corpus-assisted translation teaching, testing, and questionnaire surveys were conducted with the students in the experimental group. During the same period, traditional translation classroom teaching was conducted with the students in the control group. Finally, the same materials were selected for testing in order to compare the results between the experimental and control groups.

4.2 Training Design

The experiment lasted for four weeks. Both the experimental group and the control group received three 90-min translation class hours every two weeks (with each class hour lasting 45 min), resulting in a total of 12 experimental class hours for each group. In these 12 class hours, the teaching content was the same for both groups, but the teaching methods were completely different.

For the students in the experimental group, the teacher organized autonomous learning by utilizing basic information about corpora and establishing their own corpora on SketchEngine. The distribution of class hours was shown in Fig. 1 as follows:

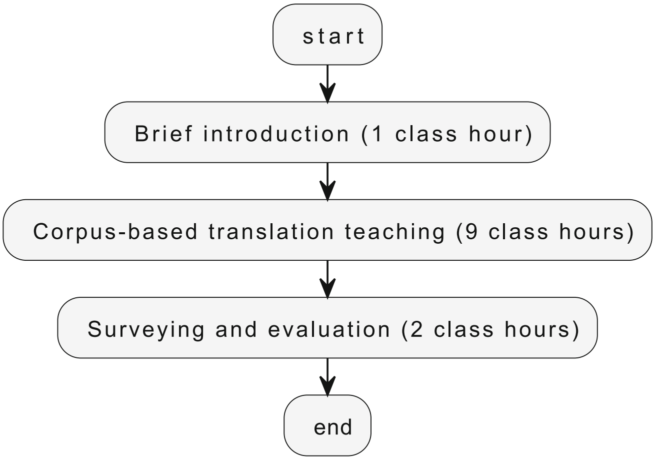


Fig. 1. Illustration of the distribution of class hours.

The specific operations for corpus-based translation teaching were as follows: The students were organized to collect Chinese political materials from the internet and create their own parallel corpus of Chinese political texts on the SketchEngine platform.

Additionally, the teacher distributed approximately 800-word Chinese political texts to the students at the beginning of the class, and the students used the parallel corpus to translate the materials. During this process, students were allowed to discuss with each other. The teacher guided students in summarizing the terminology, collocations, and sentence patterns in the materials. After completing the translation task, the teacher provided the students with the standard translation of the Chinese materials, allowing the students to autonomously check and supplement their own translations.

For the students in the control group, the teacher adopted the traditional translation teaching mode. In other words, the teacher explained the materials in detail in class, enabling the students to acquire translation skills.

After four weeks, both the experimental group and the control group of students underwent a closed-book translation test using the same test papers, which were related to the teaching content within the four weeks, in order to compare the teaching effects of the two groups.

4.3 Analysis of Pre- and Post-test Scores for Experimental and Control Groups

The final test involved translating a Chinese text of approximately 300 characters within one hour. Initially, the instructor scored the translations based on faithfulness and fluency. To ensure the accuracy of the scores, additional translation instructors were invited to provide their own evaluations. The final score for each student was determined by averaging the scores from two teachers.

Taking advantage of SPSS, the author conducted the Independent Samples T Test on the experimental group and the control group, and calculated the average scores of the two groups. The results are as follow.

As is shown in the Table 1, the average score of the experimental group is 71.278 and the average score of the control group is 77.143. Since the P value (significance) is 0.005 which is less than 0.05, we can assume that there is a significant difference between the experimental group and control in terms of overall translation level. It can be concluded that there is a significant difference in overall translation proficiency between the experimental group and the control group, with the experimental group demonstrating superior translation quality. This suggests that corpus-based translation pedagogy contributes to enhancing the effectiveness of translation instruction.

Table 1. Results of Independent Samples T Test for the Experimental Group and Control Group in Terms of Overall Language Level. The mean difference is significant at the 0.05 level.

Group	N	Mean Scores	Std. Deviation	t	Sig
Experimental Group	36	71.278	7.648	−2.921	0.005***
Control Group	35	77.143	9.217		

4.4 Questionnaire Survey Results

After the test, the students of the experimental group filled out a questionnaire about corpus-based translation pedagogy. Based on the questionnaire, the author could explore the students' attitude towards the new teaching mode as well as whether corpus-based translation pedagogy can help them to acquire translation skills better and improve their translation level.

The survey results showed that: over 60 percent of the participants believed that corpus-based teaching helped them learn translation better. The details can be listed as Fig. 2.

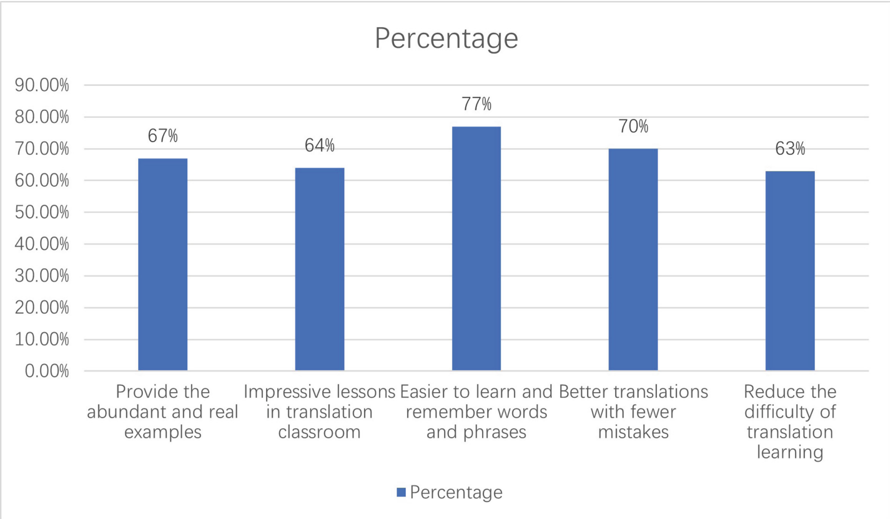


Fig. 2. Advantages of using corpora for translation assistance.

According to Fig. 2, we can see that, 67% thought that the abundant real examples provided by the corpus made it easy for them to remember what they needed to search for. 64% found the content learned in the corpus-based translation classroom impressive even after class. 77% stated that using words and phrases retrieved from the corpus made a deeper impression on them. 70% believed that using the corpus for translation made their target language more authentic and helped reduce translation errors. 63% found that using the corpus for translation reduced the difficulty of translation learning.

Additionally, when asked about future translation learning, 79% of the students believed that it is necessary to integrate corpus-based pedagogy into traditional translation teaching, and in the future, when faced with more tasks, students would rely on corpus tools for translation.

However, during the process of translation teaching, the author also identified certain disadvantages of corpus-based translation pedagogy, which include:

One challenge was the limited language variety covered by specific corpora, hindering students' ability to translate texts from diverse dialects and regional variations. Additionally, the static nature of corpora and the lack of real-time updates may affect the teaching of contemporary language usage and emerging terminology. Moreover, corpora may not always provide sufficient cultural and contextual information, necessitating the supplementation of learning with additional resources.

All in all, the corpus-based translation pedagogy presents both advantages and disadvantages, but the advantages outweigh the disadvantages. This approach enhances translation accuracy, promotes language proficiency, and fosters a systematic understanding of translation principles. However, it is important to address the limitations, such as the lack of context and domain specificity, to ensure a comprehensive and well-rounded translation education. Integrating supplementary resources and providing real-time updates can further enhance the effectiveness of corpus-based translation teaching.

5 Conclusion and Future Work

This paper presents a preliminary case study on the corpus-based translation pedagogy, aiming at investigating its effectiveness in translation teaching practices. By examining various aspects, including course design, corpus resources, teaching content, and implementation strategies, a new pedagogical model was proposed and evaluated at a local university.

The experimental group, which received parallel corpus-based translation instruction guided by constructivism and autonomous learning theory, exhibited higher average scores and improved translation accuracy compared to the control group. These results demonstrate that the integration of corpora in translation teaching positively impacts students' translation abilities and overall learning outcomes.

This study might shed light on translation teaching, which can be briefly summarized but not limited to as follows:

1) *Advocating Autonomous and Cooperative Learning*: The findings underscore the significance of fostering autonomous and cooperative learning environments in translation classrooms. Integrating computer-aided translation software and encouraging students to collaboratively discuss and solve translation challenges enhance their understanding and application of translation principles, leading to more effective learning experiences.
2) *Enhancing Learning Experience with Authentic Resources*: Providing students with authentic learning resources and advanced tools, such as computer-aided translation software, can significantly enhance their engagement and enthusiasm in translation learning. Utilizing a rich parallel corpus and specialized translation software opens up new possibilities for creating more interactive and engaging learning experiences.
3) *Recognizing the Impact of Technology*: The study emphasizes the need to explore the role of technology in corpus-based translation teaching. Understanding how different translation software and emerging technologies, such as machine translation and artificial intelligence, can be effectively integrated

into the pedagogical approach will be crucial for optimizing the learning experience in the future.

In light of these findings, future research in corpus-based translation pedagogy should explore various directions:

1) Exploring Different Types of Corpora: Researchers should investigate the application of different types of corpora in translation teaching beyond parallel corpora. Exploring the use of comparable corpora or specialized corpora will provide a comprehensive understanding of their potential effectiveness in enhancing translation skills and language proficiency.
2) Investigating the Impact on Different Language Pairs: Extending the study to examine the impact of corpus-based teaching on various language pairs will help determine the generalizability and adaptability of this pedagogical model across diverse linguistic contexts.
3) Emphasizing the Role of Technology: Future research should focus on the integration of cutting-edge technologies in corpus-based translation teaching, exploring their potential to further optimize learning outcomes and enrich the translation process.

In conclusion, the preliminary case study demonstrates the promising potential of corpus-based translation pedagogy in improving translation skills, language proficiency, and overall learning experiences. The findings highlight the importance of creating dynamic and interactive learning environments and embracing technological advancements to enhance translation education. By addressing the implications and future research directions, educators can continuously refine and advance the corpus-based translation pedagogy to benefit students and foster their proficiency in translation and language learning.

References

1. Boulton, A., Vyatkina, N.: Thirty years of data-driven learning: taking stock and charting new directions over time (2021)
2. Chen, M., Flowerdew, J., Anthony, L.: Introducing in-service english language teachers to data-driven learning for academic writing. System **87**, 102148 (2019)
3. Ellis, N.C.: Language acquisition as rational contingency learning. Appl. Linguis. **27**(1), 1–24 (2006)
4. Koehler, M., Mishra, P.: What is technological pedagogical content knowledge (tpack)? Contemp. Issues Technol. Teach. Educ. **9**(1), 60–70 (2009)
5. López-Rodríguez, C.I., Tercedor-Sánchez, M.I.: Corpora and students' autonomy in scientific and technical translation training. 1 (2008)
6. Meunier, F.: A case for constructive alignment in DDL: rethinking outcomes, practices, and assessment in (data-driven) language learning. In: Data-Driven Learning for the Next Generation, pp. 13–30. Routledge (2019)
7. O'Keeffe, A.: Data-driven learning-a call for a broader research gaze. Lang. Teach. **54**(2), 259–272 (2021)

8. Pérez-Paredes, P.: A systematic review of the uses and spread of corpora and data-driven learning in call research during 2011–2015. Comput. Assist. Lang. Learn. **35**(1–2), 36–61 (2022)
9. Pérez-Paredes, P., Sánchez-Tornel, M., Alcaraz Calero, J.M., Jiménez, P.A.: Tracking learners' actual uses of corpora: guided vs non-guided corpus consultation. Comput. Assist. Lang. Learn. **24**(3), 233–253 (2011)
10. Schaeffer-Lacroix, E.: Barriers to trainee teachers' corpus use. In: Data-Driven Learning for the Next Generation, pp. 47–64. Routledge (2019)
11. Shih, C.L., Shen, C.A.: A corpus-based investigation into student technical translation. J. Transl. Stud. 1–31 (2005)
12. Stewart, D.: Supplying native speaker intuitions or normalising translation? Translating into English as a foreign language with the British national corpus. In: From the Abstract of a Paper in the Provisional Programme of the UMIST/UCL Conference: Research Models in Translation Studies, Manchester, pp. 28–30 (2000)
13. Xuanming, L., Bin, L.: On open-style corpus-based translation teaching. Foreign Lang. Educ. **30**(6), 45–48 (2009)
14. Yoon, H., Hirvela, A.: Esl student attitudes toward corpus use in l2 writing. J. Second. Lang. Writ. **13**(4), 257–283 (2004)
15. Zanettin, F.: Corpus-based translation activities for language learners. Interpreter Transl. Trainer **3**(2), 209–224 (2009)

Investigating Student Profiles Related to Academic Learning Achievement

Yicong Liang[1], Haoran Xie[2], Di Zou[3], and Fu Lee Wang[1](✉)

[1] School of Science and Technology, Hong Kong Metropolitan University, Kowloon, Hong Kong SAR
{s1304812,pwang}@hkmu.edu.hk

[2] Department of Computing and Decision Sciences, Lingnan University, New Territories, Hong Kong SAR
hrxie@ln.edu.hk

[3] Centre for English and Additional Languages and School of Graduate Studies, Lingnan University, New Territories, Hong Kong SAR
dizou@ln.edu.hk

Abstract. In this paper, we aim to identify the key attributes from secondary school students' profiles that affect learning achievement. In particular, this work investigates and compares how demographics, school-related features and social-related features in student profiles are associated with academic success or failure in the maths final exam. The experiment is conducted on a real-world dataset, and we find that parents' education and occupation background, students' motivation and past academic records, and socializing with friends are highly associated with final learning performance. Finally, we summarize the main characteristics of students with high academic potential in secondary school.

Keywords: Educational Data Analysis · Student Performance · Learning Analytics · Student Profiles

1 Introduction

With the advancement of technologies and applications, the tremendous amount of educational data has attracted researchers to bring contributions to the field of education for analyzing students' academic performance. Online learning platforms, e.g., massive open online courses (MOOCs), have drawn significant attention from both college students and professors around the world in higher education over the past few years. However, MOOCs have been struggling with lower completion and higher dropout rate [7,14,15] compared to on-campus learning. Hence, identifying at-risk students can facilitate tutors to conduct necessary interventions, so as to retain those students in the online platforms and help them accomplish their programs [8].

Machine learning and artificial intelligence algorithms have emerged as potential tools to accomplish some learning analytics tasks, e.g., classification task of

H. Xie et al. (Eds.): ICWL 2023, LNCS 14409, pp. 39–48, 2023.
https://doi.org/10.1007/978-981-99-8385-8_4

course dropout prediction [6], and regression task of student exam performance prediction [16]. Specifically, the systems classify each student as either having a "high" or "low" risk of dropping out from the course. In addition, the system predicts the future student exam performance, i.e., the grade in the final exam.

A feasible way to help students succeed and improve the quality of education is to follow up on student performance, identify the factors affecting students' achievements. Gender is an important factor to indicate dropout in online learning environments. In MOOCs, female learners are more inclined to drop science courses while male users are more likely to give up non-science courses [6]. Additionally, there exists gender bias among learners when registering online courses, particularly in STEM and social science subjects [13].

Identifying pertinent features that influence student learning achievements is a challenging task as it heavily depends on domain expertise. In this paper, we adopt a data-analysis method to investigate what features in students' profiles are highly correlated to their final academic performance. In particular, this work focuses on investigating demographics, school-related and social-related features corresponding to academic learning performance among students from secondary school.

2 Literature Review

Identifying those students who are at risk of failing or withdrawing from their selected courses is a critical step in the process of providing them with the necessary support. Typically, interventions are made by a tutor who receives alerts of possible students' course dropout [10,16]. There are some studies that mainly focus on analyzing exam performance and aim to predict learning outcomes by solving a binary classification task, e.g., dropout or non-dropout [6], pass or fail [11]. In current online learning platforms, first assignment submission is an important indicator to identify possible at-risk students in the early stage [8]. Basically, these works build a predictive model by using various machine learning techniques, like attention-based neural network CFIN [6], XGBoost-based framework Ouroboros [8].

Another family of research works contributes to exploring relevant features or factors associated with learning performance. In the context of investigating factors in the high dropout rate in online courses, gender and friend relationships are two important factors that influence the learner's dropout probability in XuetangX [6]. Regarding the gender factor, there exists a bias in course registration in online platforms [13], where female learners prefer to take courses related to social science and male students favor STEM subjects. In addition, Aldowah et al. invited experienced instructors to assess the level of influencing factors in MOOCs [1]. The results in [1] imply that academic skills, abilities, and prior experience are essential factors related to student's decision of course dropout. Previous works [17–20] mainly focus on investigating relevant features from online learners, but this paper aims to probe profiles from students in secondary school. This work scrutinizes the related features from family, school, and social relationships and compares how these features are associated with students' learning achievement.

3 Methodology

The dataset Student Performance [3] provides student profile including demographics, school and social-related features (e.g., parent's education, past failure records, time of going out with friends), as well as grades related to the course subject. Overall, this work conducts the investigation for answering the following research questions:

RQ1 How are the learning achievements correlated with student demographics, school, and social-related features respectively?
RQ2 What are the main characteristics of students who exhibit promising academic potential in their studies?

For RQ1, this work adopts conditional probability to measure the association between the student's feature and the final academic result [13]. In particular, the feature of the student is denoted as X, and the final result is denoted as label Y. In the experiment, we find that the difference of conditional probability $p(Y = `pass'|X_{high} = 1)$[1] and $p(Y = `pass'|X_{high} = 0)$ is large, which suggests that the desire of students to go to school is an important indicator. Intuitively, students who want to pursue higher education achieve a higher pass rate than those who do not. For RQ2, this work extracts characteristics from students with high academic potential based on the following condition: If $p(Y|X) > p(Y)$, X is a salient feature. The probability $p(Y)$ reflects the student's overall distribution of success, e.g., the average pass rate of a course. The conditional probability $p(Y|X)$ reveals the specific distribution of success, for example, the pass rate of some students with certain features.

4 Experiment

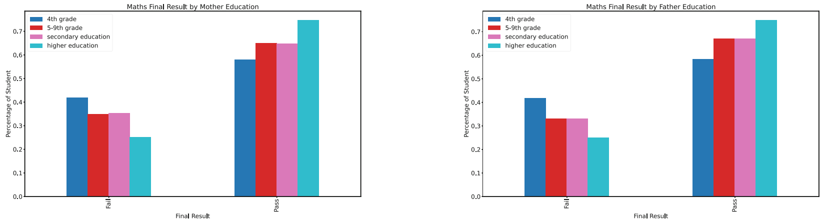
4.1 Experimental Settings

The experiment was conducted on the public dataset: Student Performance [3] in the UC Irvine machine learning repository. In this dataset, students come from secondary education from two Portuguese schools. The detailed description and statistics can be referred to [4]. The final grade, a numeric value from 0 to 20, was transferred to the binary result. Specifically, if a student obtains her final grade is greater than or equal to 10, then the corresponding result is marked as *pass*, otherwise marked as *fail*.

[1] Denote the pass rate conditioned on the student wants to take higher education in the future. ($X_{high} = 1$).

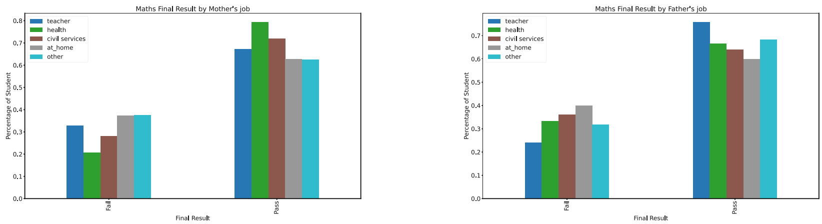
4.2 Results in Demographics

From Fig. 1, the results suggest that mothers and fathers with high education will have a positive impact on their children's academic performance. Specifically, when parents were in high education background, their children achieved a 75% pass rate. However, for those students whose mothers and fathers were with low education levels (e.g., 4th grade or below), about 42% of them failed the final exam. According to the results related to the parents' occupations shown in Fig. 2, the students whose parents had no jobs (i.e., 'at home') experienced nearly 40% fail rate. Conversely, outstanding learners came from those families whose mothers worked in hospitals or fathers were teachers. The results in Fig. 1(b) and Fig. 2(b) suggest that if the father has a higher education qualification and works as a teacher, he is able to guide his child's studies, ensuring that the child does not struggle with homework for extended periods.



(a) Mother Education (b) Father Education

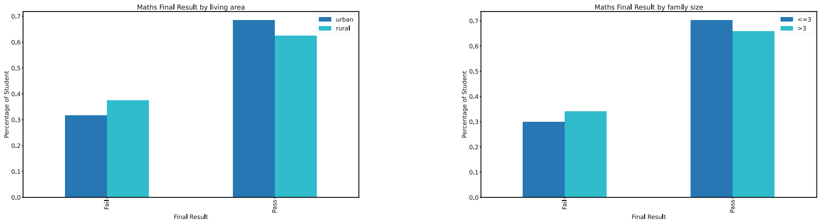
Fig. 1. Parent education background



(a) Mother's job (b) Father's job

Fig. 2. Parent occupation

There are some demographic features that are not strong indicators for learning performance, e.g., home address and family size. Urban areas are more prone to experiencing higher rates of urbanization and industrialization in comparison to rural areas. But as shown in Fig. 3(a), the living area does not have strong impact on student performance, as students living in the city center or countryside performed similarly. And students from small family or large family also achieved comparable performance, as shown in Fig. 3(b).

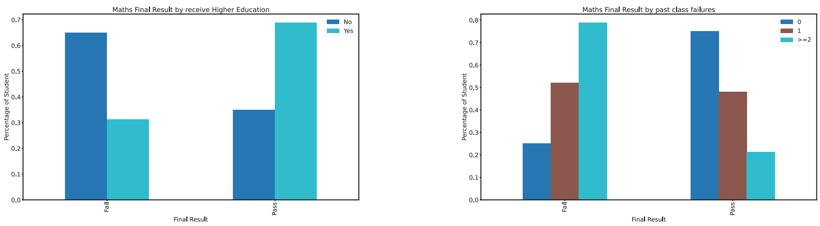


(a) Home address (b) Family size

Fig. 3. Family

4.3 Results in School-Related Features

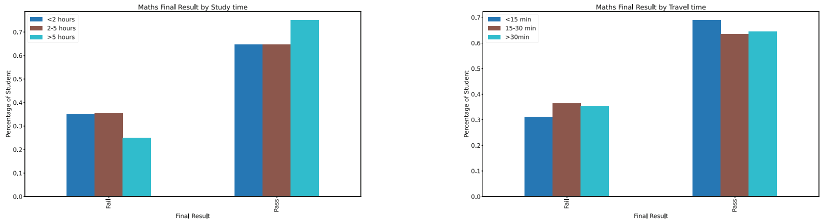
The majority of students who pass the exam demonstrate a strong desire to pursue higher education. Specifically, students who lacked the motivation to attend college had a failure rate of 65%. It is in line with the finding in [1] that motivation is a core factor for online students to learn in MOOCs. Students without failure records in the past achieved a 75% success rate in the final exam. On the contrary, students who failed more than two times would have an approximately 80% of failing the next exam. This implies that students should maintain consistent effort in their studies and strive to keep their academic records clean (Fig. 4).



(a) Willingness to take higher education (b) Failure records

Fig. 4. Student's motivation and performance in the past

From Fig. 5(a), students who spent more than five hours per week for study performed much better than those spending less than two hours each week, e.g., studying hard improves 10% pass rate in the maths final exam. But the impact from travel time is not as high as study time as shown in Fig. 5(b). Nevertheless, spending more time studying and less time traveling benefits students in achieving good grades.



(a) Study time (b) Travel time

Fig. 5. Time factor

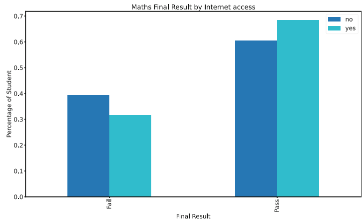
As shown in Fig. 6, the students who were able to access the Internet achieved a higher pass rate. Additionally, the students who paid extra money for maths classes performed better than those who didn't. The results suggest that learners who actively seek out learning materials and acquire more knowledge from other classes tend to achieve better results in their exams. However, according to Fig. 6(c), students who attend extracurricular activities or not performed similarly in terms of approaching pass rate, suggesting that taking extracurricular activities did not hurt the performance in the academic exam.

4.4 Results in Social-Related Features

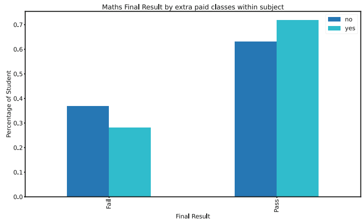
According to the results in Fig. 7, having a romantic relationship would hurt the performance in the exam. In addition, the students going out with friends too much performed worse compared to those who went out less frequently. Particularly, the pass rate from students who had no romantic relation was 10% higher than those who fell in love. Going out with friends too often dramatically degrade students learning outcome from 76% pass rate to 54%. According to the results in Fig. 8, drinking alcohol too much negatively affects student's learning performance with an increasing fail rate of 8% on both workdays and weekends. However, consuming a moderate amount of alcohol on weekends did not dramatically harm academic performance, suggesting that students can enjoy themselves and find refreshment through responsible drinking.

4.5 Main Features for Boosting Academic Potential

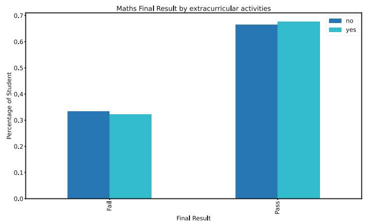
In this experiment, the overall distribution for secondary school students performing in the final exam is 67% pass rate and 33% fail rate. We selected those salient features with conditional probability $p('pass'|X) \geq 70\%$ and listed the



(a) Access to Internet

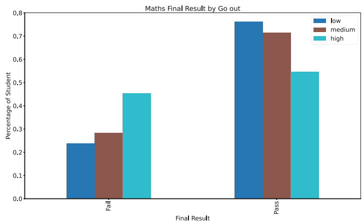


(b) Extra paid classes

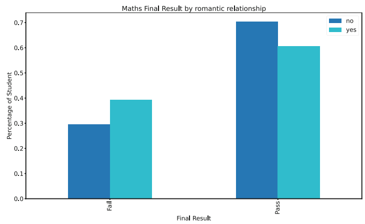


(c) Extracurricular activities

Fig. 6. After-school activity factor

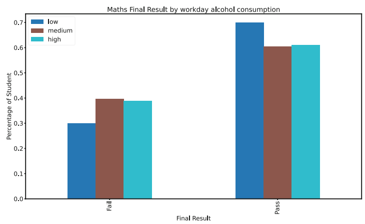


(a) Frequency of going out with friends

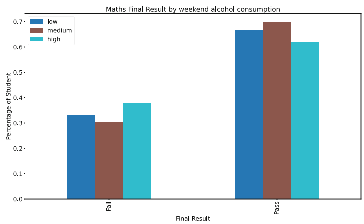


(b) Romantic relation

Fig. 7. Companion relationship



(a) Workday



(b) Weekend

Fig. 8. Alcohol consumption

top-10 results in Table 1. Parent's backgrounds including occupation and education impact positively on their children. Outstanding secondary school students originate from families where parents hold college degrees and have professional jobs. Previous studies have found that students who are more active and engage frequently with virtual learning environment system tend to have distinct grades in online learning platforms [13]. Similarly, in this investigation, secondary school students with high academic potential are expected demonstrate self-discipline, e.g., dedicating more time to studying, attending extra paid classes within specific subjects after school, and limiting unnecessary socializing time with friends.

Table 1. Top-10 features for promising academic potential in high school

Feature	Condition	Pass Rate
Mother's job	health-related	79.4%
Go out with friends	low frequency	76.2%
Father's job	teacher	75.9%
Study time per week	>5 h	75.0%
Past course failures	none	75.0%
Father education	high education	75.0%
Mother education	high education	74.8%
Attend extra paid classes	yes	71.8%
Romantic relationship	no	70.3%
Family size	≤ 3	70.2%

5 Conclusions

The recent rise in the accessibility of learning data has significantly heightened the importance and momentum of analyzing educational data to better understand and optimize the learning experience. This study conducted an investigation with a publicly accessible dataset to explore relatedness between learning achievement and student demographics, school-related features, as well as social-related features, respectively. The experiment results reveal that parents' occupations and educational backgrounds, students' individual efforts, and socializing appropriately with friends, have a positive impact on student learning outcomes. In the end, based on the experimental results, we give a summary of the main features related to those students with high academic potential studying in high school. For our future research directions, we plan to build multi-level student profiles [2,5] to capture the features at different levels and identify the students' emotions from the primitive data [9,12].

Acknowledgements. The research has been supported by IICA Project entitled "Developing language teachers' technological pedagogical content knowledge and enhancing students' language learning in virtual learning environments" (102707), the Direct Grant (DR23B2), and the Faculty Research Grants (DB23A3 and DB23B2) of Lingnan University, Hong Kong.

References

1. Aldowah, H., Al-Samarraie, H., Alzahrani, A.I., Alalwan, N.: Factors affecting student dropout in moocs: a cause and effect decision-making model. J. Comput. High. Educ. **32**, 429–454 (2020)
2. Cai, Y., Li, Q., Xie, H., Min, H.: Exploring personalized searches using tag-based user profiles and resource profiles in folksonomy. Neural Netw. **58**, 98–110 (2014)
3. Cortez, P.: Student Performance. UCI Machine Learning Repository (2014). https://doi.org/10.24432/C5TG7T
4. Cortez, P., Silva, A.M.G.: Using data mining to predict secondary school student performance. In: Proceedings of 5th Annual Future Business Technology Conference, Porto, 2008. EUROSIS-ETI (2008)
5. Du, Q., Xie, H., Cai, Y., Leung, H.f., Li, Q., Min, H., Wang, F.L.: Folksonomy-based personalized search by hybrid user profiles in multiple levels. Neurocomputing **204**, 142–152 (2016)
6. Feng, W., Tang, J., Liu, T.X.: Understanding dropouts in moocs. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 33, pp. 517–524 (2019)
7. He, J., Bailey, J., Rubinstein, B., Zhang, R.: Identifying at-risk students in massive open online courses. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 29 (2015)
8. Hlosta, M., Zdrahal, Z., Zendulka, J.: Ouroboros: early identification of at-risk students without models based on legacy data. In: Proceedings of the Seventh International Learning Analytics & Knowledge Conference, pp. 6–15 (2017)
9. Huang, X., Rao, Y., Xie, H., Wong, T.L., Wang, F.L.: Cross-domain sentiment classification via topic-related tradaboost. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 31 (2017)
10. Jayaprakash, S.M., Moody, E.W., Lauría, E.J., Regan, J.R., Baron, J.D.: Early alert of academically at-risk students: an open source analytics initiative. J. Learn. Anal. **1**(1), 6–47 (2014)
11. Kuzilek, J., Hlosta, M., Herrmannova, D., Zdrahal, Z., Vaclavek, J., Wolff, A.: Ou analyse: analysing at-risk students at the open university. Learning Analytics Review, pp. 1–16 (2015)
12. Liang, W., Xie, H., Rao, Y., Lau, R.Y., Wang, F.L.: Universal affective model for readers' emotion classification over short texts. Expert Syst. Appl. **114**, 322–333 (2018)
13. Liang, Y., Zou, D., Wang, F.L., Xie, H., Cheung, S.K.: Investigating demographics and behavioral engagement associated with online learning performance. In: International Conference on Blended Learning, pp. 124–136. Springer, Cham (2023). https://doi.org/10.1007/978-3-031-35731-2_12
14. Tang, J.K.T., Xie, H., Wong, T.-L.: A big data framework for early identification of dropout students in MOOC. In: Lam, J., Ng, K.K., Cheung, S.K.S., Wong, T.L., Li, K.C., Wang, F.L. (eds.) ICTE 2015. CCIS, vol. 559, pp. 127–132. Springer, Heidelberg (2015). https://doi.org/10.1007/978-3-662-48978-9_12

15. Wang, J., Xie, H., Wang, F.L., Lee, L.K., Au, O.T.S.: Top-n personalized recommendation with graph neural networks in moocs. Comput. Educ. Artif. Intell. **2**, 100010 (2021)
16. Wolff, A., Zdrahal, Z., Herrmannova, D., Kuzilek, J., Hlosta, M.: Developing predictive models for early detection of at-risk students on distance learning modules. In: Machine Learning and Learning Analytics Workshop at The 4th International Conference on Learning Analytics and Knowledge (LAK14). Citeseer (2014)
17. Xie, H., Zou, D., Lau, R.Y., Wang, F.L., Wong, T.L.: Generating incidental word-learning tasks via topic-based and load-based profiles. IEEE Multimed. **23**(1), 60–70 (2015)
18. Xie, H., Zou, D., Wang, F.L., Wong, T.L., Rao, Y., Wang, S.H.: Discover learning path for group users: a profile-based approach. Neurocomputing **254**, 59–70 (2017)
19. Xie, H., Zou, D., Zhang, R., Wang, M., Kwan, R.: Personalized word learning for university students: a profile-based method for e-learning systems. J. Comput. High. Educ. **31**, 273–289 (2019)
20. Zou, D., Wang, M., Xie, H., Cheng, G., Wang, F.L., Lee, L.K.: A comparative study on linguistic theories for modeling efl learners: facilitating personalized vocabulary learning via task recommendations. Interact. Learn. Environ. **29**(2), 270–282 (2021)

Online Learning Environment with Tools

Graduation Project Monitoring Platform Based on a Personalized Supervision Plan

Florina-Cătălina Anghel and Elvira Popescu(✉)

Computers and Information Technology Department, University of Craiova, Craiova, Romania
anghel.florina.w9p@student.ucv.ro, elvira.popescu@edu.ucv.ro

Abstract. Thesis supervision is a complex and important process that can directly affect the outcome of the graduation project. Poor communication between the student and the supervisor, lack of progress tracking tools and a mismatch between the preferred supervisory styles make it more likely that the project will not be finished on time or will be abandoned. Therefore, having a dedicated tool that handles all these issues could significantly contribute to the success of the graduation project. In this context, we introduce an innovative platform called Thesico, which aims to provide comprehensive support for the thesis supervision process. More specifically, the system proposes an individualized supervision plan, based on the preferences of the student and teacher with respect to support and structure. The system also facilitates the management of the project tasks (by means of Kanban boards), the intermediary assignments and the student-teacher meetings. In addition, the quality and progress of the supervision process can be assessed by means of a dedicated collaboration score computed by the platform.

Keywords: Thesis supervision · Graduation project · Supervisory style · Supervision plan · Student-teacher collaboration · Project management · Educational platform

1 Introduction

The final year / graduation project, together with its implementation process, plays an important role in the student's academic growth. Furthermore, the key to a successful Bachelor or Master thesis does not entirely consist in choosing the right project and supervisor, but it is also affected by the coordination approach.

Despite the latest technological evolution, some universities do not use a specialized tool for the thesis supervision process. Without an application that can be used for thesis completion monitoring, the process is based only on the supervisor's experience and preferences [10]. Furthermore, a survey was conducted in [1], which highlighted that the limitations imposed on student-teacher communication represent the main factor that can lead to the abandonment/delay of the thesis completion. A similar idea was also presented in [8], where the authors pinpointed that collaboration is affected by the parties' availability. In [1], it was also noted that no standard methodology is used for the thesis supervision process and the communication between the supervisor and the

H. Xie et al. (Eds.): ICWL 2023, LNCS 14409, pp. 51–61, 2023.
https://doi.org/10.1007/978-981-99-8385-8_5

student is handled using various ad hoc means, e.g. email, phone calls, SMS. While this approach is flexible since it is chosen by the student-teacher pair, it can also have limitations or cause problems like losing track of the implemented work and/or of the provided feedback, being hard to manage and review the current status of the thesis [1].

Another factor that directly affects the quality of the collaboration between the student and the teacher is represented by the number of supervisees for a single supervisor, which without a specialized platform, can be very tedious for the instructor [5].

Since the student-teacher collaboration depends on the compatibility between the involved parties, their preferences should be taken into account. Hence, the supervision process must be suitable for both the student and the teacher preferences and schedules.

In addition, Romdhani et al. pinpointed that the student learning experience can be improved by using a specialized tool [10]. Starting from these ideas, we propose a dedicated platform, called Thesico (**Thesi**s **co**ordination), which is a web application that can be used to handle the student-teacher collaboration process and project management while working on the thesis. By means of Thesico, the users will be able to follow a supervision plan (suggested by the system based on the preferences of the parties or custom, proposed by the supervisor), and to manage the meetings, assignments, and project tasks. Furthermore, they will be able to evaluate the collaboration process by visualizing various statistics (including a collaboration score).

The rest of paper is structured as follows. The next section provides an overview of the related work which includes similar systems and the theoretical aspects that have been considered while implementing the Thesico platform. The third section presents the mechanism and workflow of the Thesico system, while the fourth section illustrates its main functionalities. Lastly, the fifth section includes some conclusions and future research directions.

2 Related Work

Two of the most important steps that can assure the success of a graduation project are: i) choosing a project theme and supervisor that match the student's needs and ii) smoothly handling the interactions with the supervisor to discuss the current state of the project, the next tasks, the challenges encountered and any other organizational elements that are needed in the project implementation phase.

Regarding the first step, studies show that a positive relationship between the student and the supervisor and a good compatibility between them with respect to personality and supervisory style can lead to increased engagement, performance and well-being [6, 7, 11]. In this context, we already implemented a platform called SPA (Smart Project Allocation) which aims at finding the best fit for the student based on several factors (supervisory style and personality of the teacher, domain and complexity level of the project) [2]. In addition, SPA provides a way to dynamically allocate the thesis projects based on the student-teacher and student-project compatibilities, in a fair and transparent manner. We therefore aim to integrate the SPA system with our proposed platform, in order to handle this first matching and allocation step.

Regarding the second step (i.e., project monitoring and coordination), Almeatani et al. suggest that in general the resources provided by the university (free accounts for different project management tools or access to resources) are not enough for the student and the supervisor since they can lose track of the discussed aspects (status, feedback, planned meetings) and their schedules also affect their interaction [1]. Since the supervision process is very complex, it needs to be supported by a specialized tool that will handle the management of each component of this process; hence, several systems have been proposed in the literature for this purpose.

One such system is Git4School, introduced in [9], which includes a dashboard with statistics regarding the students' GitHub[1] activity. For example, the statistics are built based on the homework status (if it was implemented without help from other students or if it was implemented before the teacher's solution has been published). While this system aims at providing an overview of the student's activity based on the homework assigned by the teacher, the authors conclude that the usage of a specialized tool improves the supervision process by providing a real-time overview of the learners' activity which can be used to identify the students that need more support from the teacher [9].

Another similar platform, called TSS (Thesis Supervision System), is proposed in [1] and it can be used to handle two types of interactions: the supervisor-supervisee interaction and the student-student communication. Using TSS, the supervisors can provide feedback and both students and teachers can schedule meetings. Furthermore, the supervisees can manage their project tasks, under the coordinator's supervision and they can also interact with other students that have projects from the same domain and they can share information [1].

A related project supervision system is presented in [10], which handles the project allocation part but also provides the possibility to manage the tasks and milestones that should be validated by the teacher. In addition, more generic project management platforms such as Jira[2] could also be used to monitor the tasks of a student project, but without dedicated educational support features.

Our goal is to integrate the most useful functionalities from the above systems, but also take into account the students' and teachers' preferences with respect to supervisory style. Implementing a personalized supervision plan is an innovative feature of our Thesico platform, which builds on the user profiles created by the SPA system. More specifically, the supervisory style consists in the principles that are present in the teacher-student collaboration (supervisor-supervisee relationship) [3]. We follow the conceptual model proposed by Gatfield [4], in which the supervisory styles are characterized by two factors, structure and support; four supervisory styles are thus defined: *Laissez-faire, Directorial, Pastoral,* and *Contractual*.

Firstly, the *Laissez-faire* style is defined by a low level of structure and support (i.e., the supervisor provides limited task guidance or personal interaction). Secondly, the *Pastoral* style is described as an approach with a low level of structure, but a high level of support (i.e., the supervisor offers personal care but limited task-driven directions). Thirdly, the *Directorial* style is represented by a high level of structure but a low level of support (i.e., the supervisor tends to focus on the project tasks, but less on the personal interaction). Lastly, the *Contractual* style is defined by a high level of structure and support (i.e., the supervisor provides both project task direction and personal care) [4].

[1] https://github.com.

[2] https://www.atlassian.com/software/jira.

Starting from the students' and teachers' preferred supervisory style, Thesico platform proposes an individualized supervision plan, as described in the next section.

3 Thesico Prototype

3.1 Designing an Approach for Individualized Project Supervision

Thesico platform was designed in order to support and facilitate the thesis supervision process, by providing a standard workflow for project tasks, assignments, guidance meetings and milestone meetings. The project management issues, as well as the communication and collaboration between the supervisor and the student are thus handled in one comprehensive platform.

With respect to project management, a Kanban board[3] is included, which provides a visual representation of the work in progress, with cards that represent tasks, structured in columns that represent the work status. Regarding the supervision process, the concept of a custom plan is introduced, which aims to take into account the preferred supervisory style of the student and teacher. As mentioned in the previous section, each supervisory style is defined by two components: structure and support. In Thesico, the structure component is based on two platform features: the assignments and the milestone meetings. The assignments are used to handle the thesis documentation completion process (intermediary and final versions), while the milestone meetings are oriented on specific project tasks, where the student presents the progress of the project. On the other hand, the support component is represented by the guidance meetings between the student and the teacher, which can be used for general discussions, providing feedback and organizational issues. A predefined supervision plan is proposed by the system for each of the four supervisory styles (Laissez-faire, Directorial, Pastoral, and Contractual), which can be further individualized by the teacher according to the specific needs of the students.

More specifically, the Thesico workflow is as follows. First of all, the administrator configures various system settings, such as the number of milestone meetings, the number of assignments and guidance meetings for each supervisory style, the penalty for delayed assignments, the weights used to compute the collaboration score (more details will be presented in the following subsection). Next, the administrator imports the student and teacher data from the SPA platform [2], which was previously used for matching students with supervisors and dynamically allocating the graduation projects.

Subsequently, the actual thesis development and supervision process starts. The teacher can customize the predefined supervision plan, schedule meetings, create assignments, add project tasks, grade assignments and milestone meetings as well as the final thesis. The student works on their project, submits assignments, and updates the tasks status on the Kanban board. Both the student and the teacher are asked to assess the usefulness of the guidance meetings by providing a rating after each such meeting, which will be further used for computing a collaboration score, as described in the following subsection.

[3] https://en.wikipedia.org/wiki/kanban_board.

More details regarding the Thesico functionalities are presented in Sect. 4. With respect to implementation, the platform was built as a RESTful application with a client-server structure. The frontend was developed in TypeScript using the Vue.js framework[4]. The backend is based on a REST API implemented using the PHP programming language and the API Platform framework[5], which internally uses the ADR pattern (Action Domain Responder)[6].

3.2 Assessing the Student-Supervisor Collaboration

The Thesico system provides the possibility to evaluate the quality and progress of the supervision process by computing an *overall collaboration score* based on the following components: structure, support and user rating.

Firstly, the *structure score* aims to assess the completion rate for the two structure features: the assignments and the milestone meetings. The *assignments score* indicates whether the work was submitted on time by the student; for each assignment, the score is computed according to the following formula:

$$Sa_i = \begin{cases} 100,\ if\ assignment\ i\ was\ turned\ in\ on\ time \\ 100 - p * d,\ if\ assignment\ i\ was\ turned\ in\ late \\ 0,\ if\ assignment\ i\ was\ not\ turned\ in \end{cases}$$

where p represents the penalty for delayed assignments (per day) and d represents the number of days late (note that if the result is negative, i.e. the assignment was turned in very late, then the score is set to 0). An *average score for all the assignments* is subsequently computed as follows:

$$Sa = \frac{\sum_{i=1}^{n} Sa_i}{n}$$

The *milestone meetings score* measures the completion rate for the predefined milestone meetings, as follows:

$$Smm = \frac{number\ of\ completed\ milestone\ meetings}{set\ number\ of\ milestone\ meetings\ that\ need\ to\ be\ completed} * 100$$

Thus, the *structure score* ($Sstr$) is computed as the average of the two scores above (Sa and Smm).

Secondly, the *support score* ($Ssup$) measures the completion rate for the predefined guidance meetings, as follows:

$$Ssup = \frac{number\ of\ completed\ guidance\ meetings}{set\ number\ of\ guidance\ meetings\ that\ need\ to\ be\ completed} * 100$$

Thirdly, the *rating score* (Sr) is computed based on the ratings provided by the teacher and the student with respect to the perceived usefulness of the guidance meetings.

[4] https://vuejs.org.

[5] https://api-platform.com.

[6] https://github.com/pmjones/adr.

Finally, the *overall collaboration score* (*Sc*) represents the weighted average of the three scores above:

$$Sc = Sstr * w\_str + Ssup * w\_sup + Sr * w\_r$$

where *w_str*, *w_sup* and *w_r* are the weights associated to each score (structure, support and rating score respectively), which can be configured by the administrator.

4 Illustrating Thesico Functionalities

As mentioned before, the Thesico platform has three user roles (administrator, teacher / supervisor and student / supervisee). In the following three subsections we present the perspective of each role respectively, illustrating the main functionalities provided by the system.

4.1 Administrator Perspective

The system administrator has the main role of retrieving data from the SPA platform (which was used in a previous step for allocating thesis projects based on student-teacher compatibility factors) and importing it into Thesico. In addition, the administrator is able to configure various system settings, based on the specific requirements of the university program:

- *General settings* – the upper limit for the number of milestone meetings and the penalty applied to late assignments
- *Supervisory styles settings* – the recommended number of assignments and guidance meetings for each supervisory style
- *Weight settings* – the predefined weights for the supervisor's ratings as well as for the structure, support and rating scores, used for computing the overall collaboration score.

4.2 Teacher Perspective

The teacher (supervisor) is provided with the following main functionalities in Thesico:

- *Dashboard* – a page used to view general information (the grades and the collaboration scores of the supervised students) and a calendar with the guidance and milestone meetings for all supervised projects. This functionality is presented in Fig. 1.
- *Student management* – module used to view, search and order the supervised students. For each student, the teacher can define the supervision plan, see the project description, provide a final grade for the thesis and access the project module.
- *Project statistics* – the teacher can see various details regarding each project, e.g.: percentage of completed guidance meetings / milestone meetings / assignments, collaboration scores, charts with the status of the project tasks (as depicted in Figs. 2 and 3).

- *Meetings calendar* – the teacher can schedule meetings (guidance or milestone), edit / cancel a meeting, mark it as missed or completed. Furthermore, if a milestone meeting is completed, the teacher can provide a grade (to assess the student's progress) and if a guidance meeting is completed the teacher can provide a rating (to assess the usefulness / satisfaction with the meeting).

Fig. 1. Thesico - Supervisor dashboard

- *Assignments calendar* – the teacher can create / edit / delete an assignment, view the documents uploaded by the student (if available) and grade a submitted assignment.
- *Project tasks / issues* – the teacher can access a page where they can view the tasks of the project using two different components: a table and a Kanban board. The issue types are taken from Jira[7] (e.g., Epic, Story, Task, Bug, and Subtask type); the issues

[7] https://support.atlassian.com/jira-cloud-administration/docs/what-are-issue-types.

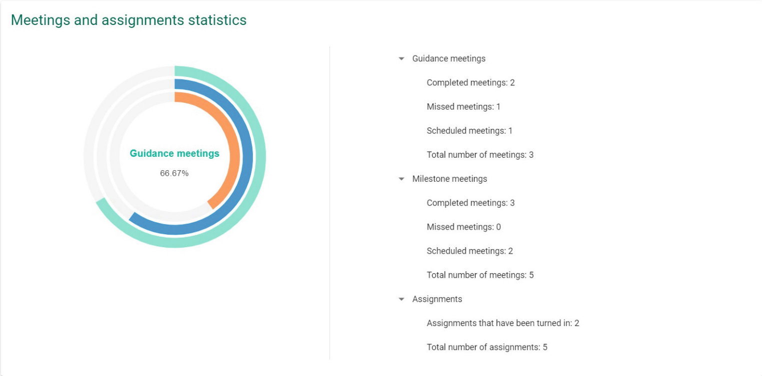


Fig. 2. Thesico – Project statistics: meetings and assignments status

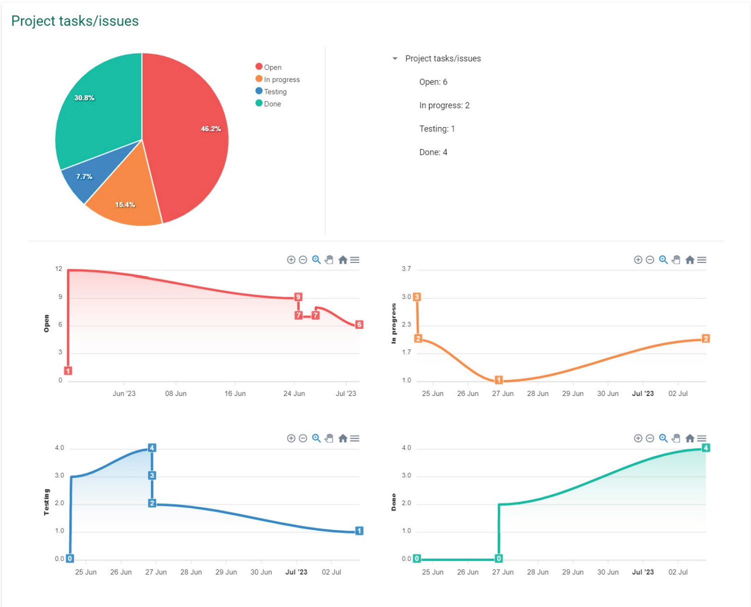


Fig. 3. Thesico – Project statistics: tasks status

that have the type set to Epic cannot have parent issues and are used to organize bigger pieces of work that need to be done. Thus, these items will not be displayed in the Kanban board but in a separate table and the user will be able to order and search them. The remaining types of issues will be displayed on the Kanban board for a better user experience when updating the statuses. In addition, the teacher can also access a dedicated page for each task, which contains the following data:

o General information about the task - title, description, related tasks.

- o The status of the task - the issue type, the current status, the creation date, and the date when the last update was performed. The teacher can also edit the details of each task.
- o The attachments component – which includes the documents that have been uploaded for that task (if any); the teacher can also add new attachments or delete them.
- o A timeline component – which is used to show a history with the status updates that have been performed for the current task.

4.3 Student Perspective

The student (supervisee) is provided with the following main functionalities in Thesico:

- *Dashboard* – a page used to view the collaboration score and the meetings / assignments that are scheduled for the following week.
- *Project statistics* – the student can see various details regarding their project, including charts with completed meetings / assignments and tasks status (similar with the teacher perspective illustrated in Figs. 2 and 3).
- *Meetings calendar* – the student can view the meetings that are scheduled, together with various details, such as description, date, status (‘Proposed’, ‘Completed’, ‘Missed’), grade (if provided by the teacher for a milestone meeting). Furthermore, the student can give a rating for a completed guidance meeting, in order to evaluate the usefulness / satisfaction with that meeting. This functionality is presented in Fig. 4.
- *Assignments calendar* – the student can see the assignments that have been created and for each assignment they can upload / remove files and check the grade provided by the teacher.
- *Project tasks / issues* – this module is very similar to the one presented for the teacher in the previous subsection. In addition to the actions that can be performed by the teacher, the student can also update the status of each task and order them using the Kanban board. A part of this functionality is illustrated in Fig. 5.

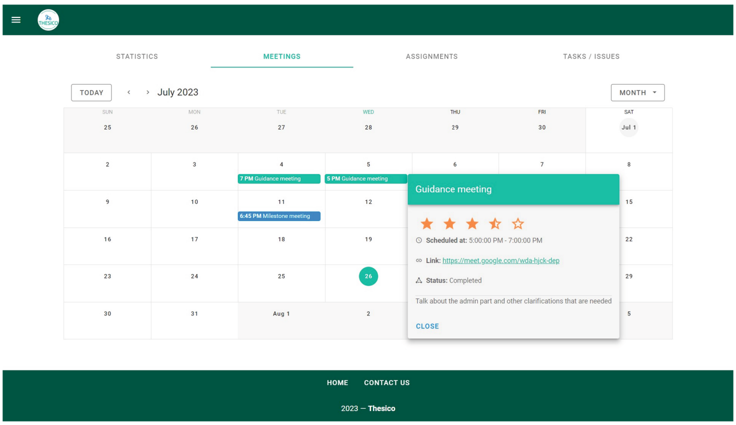


Fig. 4. Thesico – Meetings calendar

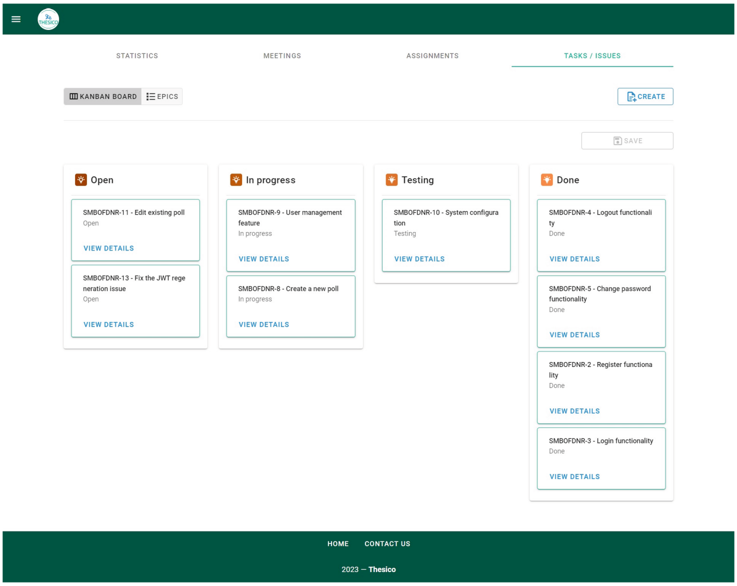


Fig. 5. Thesico – Project tasks / issues (Kanban board)

5 Conclusion

We proposed an approach for managing the coordination process of a graduation project based on a personalized supervision plan, which takes into account the preferred supervisory style of the student and the teacher. We designed and implemented a dedicated platform, called Thesico, which provides support for the thesis supervision process. By means of the system, the teacher can schedule meetings, create and grade intermediary assignments, follow the progress of the project and visualize various statistics. The student can manage the project tasks and receive feedback and guidance from the teacher throughout the process.

As future work, the system could be extended by integrating the GitHub API[8] to gather more information about the student's activity based on the performed commits on GitHub. The communication and collaboration between the student and the teacher could also be further supported by implementing a chat, as well as a wiki page for project related resources and guidelines. We also aim to use the Thesico system in real world settings, in order to assess the usefulness of the personalized supervisory plan and gauge students' and teachers' satisfaction with the platform.

References

1. Almeatani, M., Alotaibi, H., Alasmari, E., Meccawy, M., Alghamdi, B.: Thesis supervision mobile system for enhancing student-supervisor communication. Int. J. Interact. Mobile Technol. (iJIM) **13**(06), 4–14 (2019)

[8] https://docs.github.com/en/rest.

2. Anghel, F.C., Popescu, E.: Dynamic graduation project allocation based on student-teacher profile compatibility. In: Proceedings ICALT 2022 (22nd IEEE International Conference on Advanced Learning Technologies), pp. 56–60, IEEE (2022)
3. Boehe, D.M.: Supervisory styles: a contingency framework. Stud. High. Educ. **41**(3), 399–414 (2016)
4. Gatfield, T.: An Investigation into PhD supervisory management styles: development of a dynamic conceptual model and its managerial implications. J. High. Educ. Policy Manag. **27**(3), 311–325 (2005)
5. Kar, S.A.C., Fadzal, N., Ismail, S.I., Abdullah, R., Omar, H.M.: Integrated supervision and evaluation system for final year project. In: Proceedings SCOReD 2017 (15th Student Conference on Research and Development), pp. 198–203, IEEE (2017)
6. Liang, W., Liu, S., Zhao, C.: Impact of student-supervisor relationship on postgraduate students' subjective well-being: a study based on longitudinal data in China. High. Educ. **82**, 273–305 (2021)
7. Pancorbo, G., Decuyper, M., Kim, L.E., Laros, J.A., Abrahams, L., De Fruyt, F.: A teacher like me? different approaches to examining personality similarity between teachers and students. Eur. J. Pers. **36**(5), 771–786 (2022)
8. Putra, J., Yuliana, Sukiman, Waisen, Awan, Benny: Data management system for thesis monitoring at STMIK IBBI Using B-Model. In: Proceedings MECnIT 2020 (3rd International Conference on Mechanical, Electronics, Computer, and Industrial Technology), pp. 365–369, IEEE (2020)
9. Raclet, J.-B., Silvestre, F.: Git4School: A Dashboard for Supporting Teacher Interventions in Software Engineering Courses. In: Alario-Hoyos, C., Rodríguez-Triana, M.J., Scheffel, M., Arnedillo-Sánchez, I., Dennerlein, S.M. (eds.) EC-TEL 2020. LNCS, vol. 12315, pp. 392–397. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-57717-9_33
10. Romdhani, I., Tawse, M., Habibullah, S.: Student project performance management system for effective final year and dissertation projects supervision. In: Proceedings LICE 2011 (London International Conference on Education) (2011)
11. Zhang, M., Sun, J., Ma, J., Wu, T., Liu, Z.: A personality matching-aided approach for supervisor recommendation (research-in-progress). In: Proceedings HICSS 2016 (49th Hawaii International Conference on System Sciences), pp. 678–687, IEEE (2016)

An Online Tutoring and Assessment System for Teaching Relational Algebra in Database Classes

Hasan M. Jamil(✉), Kallol Naha, and Farjahan R. Shawon

University of Idaho, Moscow, ID 83844, USA
jamil@uidaho.edu, {naha7197,shaw0901}@vandals.uidaho.edu

Abstract. While there are several online tools to practice SQL, except for single tool to practice relational algebra, there practically are no system for teaching, tutoring or assessing relational query language assignments for database classes. In this paper, we introduce *ReliQ*, an online tutoring and assessment system for teaching relational query languages to database students. ReliQ supports a number of features for convenient management of assignments and tests for both practice (in tutoring mode) and authentic assessment (in testing mode). It is capable of grading assignments autonomously and generating useful hints for an effective, enriching and unparalleled eLearning experience.

Keywords: Intelligent tutoring · authentic assessment · eLearning · self-paced learning · user interface · relational algebra · database classes

1 Introduction

As delivery and management of traditional courses increasingly go online, the demands for more sophisticated learning and administrative tools are also growing. Two of the advantages online platforms promise are self-paced learning [24] and scalability [1]. The absence of a physical classroom in online learning also necessitates, among many other support systems, a robust tutoring [9] and assessment [18] system support for effective eLearning. For database teaching in particular, several decoupled tutoring systems were designed to cover topics of a first database class with varying degrees of focus and sophistication. The absence of a comprehensive database teaching, tutoring and assessment system leaves open the opportunity for developing one.

The relational algebra query language tutoring and assessment system, *ReliQ*, we present in this paper is one of the subsystems of a larger and comprehensive database tutoring and assessment system called *Project 360*. Project 360 is designed to include four inter-connected and integrated tutoring systems – Conceptual Database Design (*CoDD*), Visual SQL (*ViSQL* [11]), Relational Algebra Query Language (*ReliQ*), and Normalization and Database Design Theory (*NoDD* [7,8]). In the remainder of this article, we discuss how ReliQ approaches authentic assessment and tutoring of database courses using relational algebra querying.

H. Xie et al. (Eds.): ICWL 2023, LNCS 14409, pp. 62–76, 2023.
https://doi.org/10.1007/978-981-99-8385-8_6

2 Related Research

A recent system for conceptual database design, ERDPlus [10], was introduced. While it allows several sophisticated features to help design and understand ER diagrams and their relationships with SQL schema, it does not particularly support feedback generation or assessment. It is also not part of a tutoring system designed for class management as we discuss in this paper. Similarly, SQL tutoring systems [5,12,14], and normalization tools [22] are not comprehensive systems though they support simple problem solving features toward online tutoring. Of particular interest are the relational algebra tutoring systems RelaX [13] and a cognate system called Alloy [2] for discrete math lab exercises.

RelaX is the closest system that compares well with ReliQ that actually allows execution of relational algebra expressions. It supports most of the relational algebra expressions including several extended operators. Among the contemporary relational algebra tutoring systems, such as radb [23], Relational [21], and IRA [17], RelaX probably is the most advanced and widely used. However, this system is not suitable for a full-fledged tutoring system such as ReliQ. In particular, it does not support assignment posting, tests, or grading. Nor can it be made part of a more elaborate system without extensive redesign even though the source code is made available online. Similar observations apply to radb, Relational and IRA as well.

In contrast, ReliQ is a complete and comprehensive tutoring and assessment system. Instructors are able to design courses around ReliQ, and manage the delivery and conduct of the course in full. It supports a comprehensive set of features traditionally supported by learning management systems (LMS), and more. In particular, relational algebra questions can be asked in assignments and tests, students can be allowed to test their queries before submission, and the tests can be graded fully autonomously. A more sophisticated feedback and query explanation option for ReliQ is being developed along the lines of RATest [15].

3 Relational Algebra Assignments and Tests

A traditional relational algebra assignment usually has a description of a database scheme, a set of table schemes, and a number of queries over the database scheme written in a natural language such as English. A student's task is to implement or translate those natural language queries into relational algebra expressions that when executed will return a response semantically consistent with the queries written in English. For example, the query below

Example 1. List all cats which like to eat the most expensive Purina brand foods.

over the database scheme below where the primary keys are underlined in each table scheme.

Pets(<u>PetID</u>, Name, City, Zip, Age, PetType)
Likes(<u>PetID, Year</u>, FoodID)
Foods(<u>FoodID</u>, Name, Brand, Price)

In ReliQ, in order to compute this query using relational algebra and be graded, a student must write an expression similar (technically speaking, semantically equivalent) to the reference query provided by the instructor. However, for now, let us assume that the student wrote the query below as her answer.

$$R_1\colon\ \Pi_{Name}(\sigma_{PetType='Cat'}(Pets \bowtie Likes \bowtie \Pi_{FoodID,max(Price)\ as\ mPrice}(\sigma_{Brand='Purina'}(Foods))))$$

A more convenient and structured way of composing the query uses extended relational algebra operators such as assignment and renaming as follows.

$$R_2\colon\ Temp \leftarrow \rho_{tmp(FoodID,mPrice)}(\Pi_{FoodID,max(Price)}(\sigma_{Brand='Purina'}(Foods)))$$
$$\Pi_{Name}(\sigma_{PetType='Cat'}(Pets \bowtie Likes \bowtie Temp)$$

3.1 Question Answering in ReliQ

ReliQ supports a query editor using which users are able to construct relational algebra expressions. Figure 1 shows the ReliQ editor for query construction and execution over a target database. It has three main parts: the operator bank at the top, database table schemes to the right, and the editing pane below the operator bank. Users can choose operators by clicking on the operator icons for ReliQ to insert it in the editor pane at the selected location. Operator arguments or qualifiers (e.g., selection conditions or projection lists) can be inserted as subscripts, or on the same level of the operator, so long the syntax is semantically equivalent, i.e., ReliQ is subscript agnostic.

Fig 1 also shows ReliQ's functional components in an active editing session in which the query R_1 has been constructed and executed against a sample data, e.g., the Pet database. The result of the execution is shown as a table below the editor pane. The displayed table also has column sorting, pagination and page size selection options for convenient visualization. For validation purposes, users are also able to inspect the base table instances just by clicking on the table names at the right segment of the interface. Options to display the standard SQL and MySQL SQL equivalent of the queries in the editor pane are also available as references.

3.2 ReliQ Query Execution

ReliQ query execution is by translation of algebra queries into SQL. In ReliQ, we use MySQL version 8.0.33 on Linux. Since MySQL does not always support or follow standard SQL, ReliQ to SQL translation is MySQL specific, and follows MySQL syntax specific mapping rules. In this section, we present an overview of ReliQ expression to SQL translation procedures, and discuss how student queries are assessed against an instructor provided reference query.

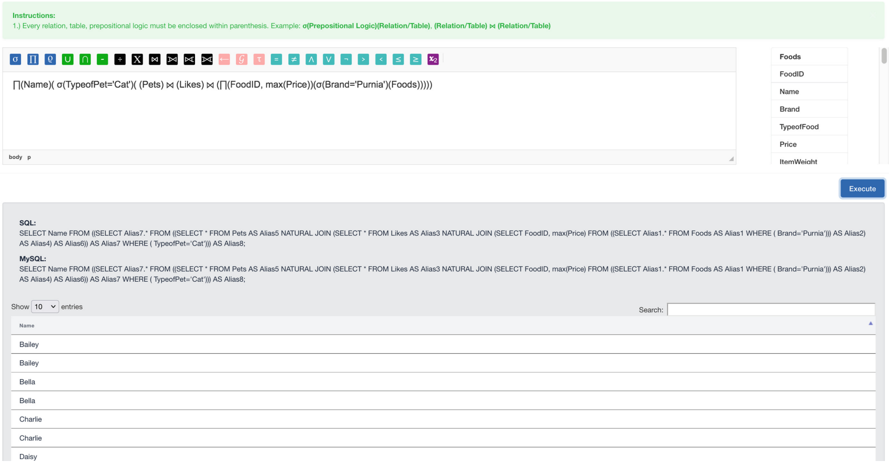


Fig. 1. ReliQ IDE and editor.

ReliQ to SQL Mapping. Query translation has been leveraged as a less costly implementation strategy in numerous applications [16,20,25] relying on its demonstrated strengths and usefulness [3,6]. The key to assigning translational semantics to a query language is to ensure that the semantic view of the source language is preserved in the target language, and if an inverse mapping to the source language is applied, there will be no "loss". In ReliQ too, we exploit this implementation strategy and translate ReliQ queries into SQL, MySQL version of SQL to be exact, for execution. In the current edition of ReliQ, our focus is on correctness of translation, and a straightforward implementation without a serious focus on efficiency concerns of the back-end query processing costs in MySQL.

Query Classes For the purpose of translation, we recognize that only set operations in ReliQ can be translated directly in SQL. For example, $\Pi_{PetID}(Pets) - \Pi_{PetID}(Likes)$ can implemented as the SQL query

(SELECT PetID FROM *Pets*)
EXCEPT
(SELECT PetID FROM *Likes*);

to discover pets which have no specific food they like to eat. In this query, two subqueries are connected as operands of the EXCEPT operator. In contrast, the query $\Pi_{PetID}(Pets \bowtie Likes)$ cannot be implemented as the SQL query below even though this query too is similar in structure and spirit to the query above.

(SELECT PetID FROM *Pets*)
NATURAL JOIN
(SELECT PetID FROM *Likes*);

Instead, this ReliQ query can be implemented as the following SQL query.

```
SELECT *
FROM (SELECT PetID FROM Pets)
         NATURAL JOIN
        (SELECT PetID
        FROM Likes);
```

Or, more directly as

```
SELECT PetID
FROM Pets NATURAL JOIN Likes;
```

Though both versions appear simple, and the second version simpler yet, the appearances are deceiving. The first version is a direct translation of the ReliQ query that exploits the knowledge that the SQL equivalent of the entire ReliQ query must reside inside a FROM clause, and not much else. In contrast, the second query though more intuitive and cleaner, it requires extensive analysis of the ReliQ query to understand the query sufficiently enough to construct an equivalent SQL version, which in some cases could turn very complicated. We, therefore, chose the first approach for the sake of simplicity even though it is computationally more expensive than the latter.

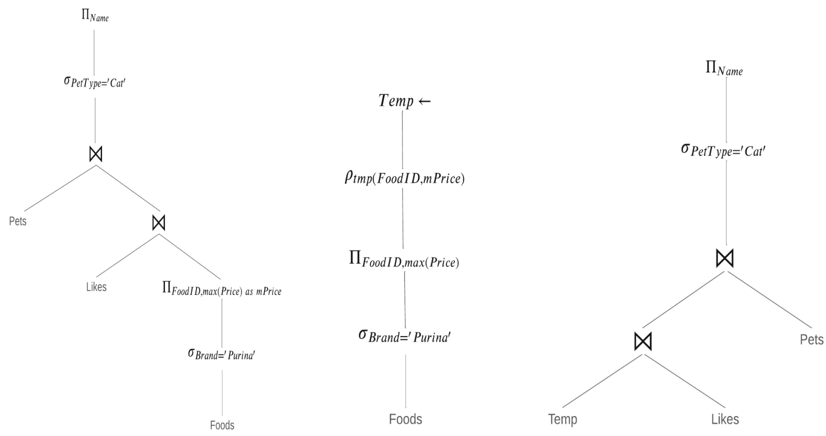
Mapping Algorithms The mapping algorithm we have developed are case by case, extensive and too elaborate to include in this article. Instead, we present the essence of the steps involved in most of these mapping algorithms using the Ex 1 and discuss the general and simple structure they follow on intuitive grounds below.

We first create the query tree from the expressions. Then we proceed to construct SQL expressions step wise from the leaf nodes to the root. The process involves recognizing unary operations (selection, projection, rename, assignment, etc.), and in expression binary operations (joins, Cartesian product, etc.) and open binary operations (union. intersection, difference), and constructing SQL expressions according to the principles discussed above in this section. Special considerations are given to special operations such as division, group by and ordering.

Figures 2(a) and 2(b) show the query trees corresponding to the ReliQ queries R_1 and R_2 respectively in which the leaf nodes are all tables. While they compute the same query and identical responses in two very distinct different ways, we choose to explain the mapping process using the query tree for R_2 in Fig. 2(b) since it covers more operators than R_1. The steps we follow to convert the left query tree of R_2 into an SQL query can be summarized as follows.

Step 1 Identify leaf node *Foods*. Construct expression e as the clause FROM Foods.

Step 2 Recognize unary operation selection. Update expression e as FROM *Foods* WHERE Brand='Purina'.



(a) Query tree for R_1. (b) Query tree for R_2.

Fig. 2. Query trees.

Step 3 Recognize unary projection operation. Update expression e as SELECT FoodID, max(Price) FROM *Foods* WHERE Brand='Purina'.
Step 4 Recognize unary operation renaming. Update e as SELECT * FROM (SELECT FoodID, max(Price) as mPrice FROM *Foods* WHERE Brand='Purina') as tmp.
Step 5 Recognize root assignment node. Create final expression e as

CREATE VIEW *Temp* as
SELECT *
FROM (SELECT FoodID, max(Price) as mPrice
FROM *Foods*
WHERE Brand='Purina') as tmp;

Finally, the right query tree in Fig. 2(b) is interpreted as the SQL query

SELECT Name
FROM *Temp* NATURAL JOIN *Likes* NATURAL JOIN *Pets*
WHERE PetType='Cat';

It does so by recognizing that *Temp* is a virtual table and needs to be computed prior to executing the query above. Also, this time, it starts with the expression FROM *Temp* NATURAL JOIN *Likes* since JOINs are in expression binary operators, and builds up the query above following steps we have outlined above. In contrast, if we were to create an SQL query for the query tree in Fig. 2(a), the query will be as follows.

SELECT Name
FROM (SELECT FoodID, max(Price) as mPrice

```
        FROM Foods
        WHERE Brand='Purina') NATURAL JOIN Likes
            NATURAL JOIN Pets
    WHERE PetType='Cat'
```

Automatic Grading of Student Assignments. While translating student ReliQ expressions into SQL is not too complex a technical challenge, grading them properly somewhat is. Recall that Ex 1 expects the list of cats that eat the most expensive Purina foods and assume that the instructor wanted the name of the cats along with the PetID or Zip to distinguish between the cats with identical names. Then the reference query she wrote is as follows:

$$R_r\colon\ \Pi_{PetID,Name}(\sigma_{PetType='Cat'}(Pets \bowtie Likes \bowtie (\Pi_{FoodID}(\sigma_{Brand='Purina'}(Foods)) \bowtie \Pi_{max(Price)\ as\ Price}(\sigma_{Brand='Purina'}(Foods)))))$$

Incidentally, the reference query R_r is the correct expression for the query in Ex 1, and therefore the student queries R_1 and R_2 are incorrect[1]. The computational challenge here is that how do we test equivalence of the two queries, i.e., $R_r \stackrel{s}{\equiv} R_1$ or $R_r \stackrel{s}{\equiv} R_2$ (here, $\stackrel{s}{\equiv}$ denotes semantic equivalence). Semantic equivalence ($\stackrel{s}{\equiv}$) technically entails $|R_1| = |R_2| = |R_r|$, and $\forall x(x \in R_1 \Leftrightarrow x \in R_r)$ or $\forall x(x \in R_2 \Leftrightarrow x \in R_r)$.

The queries R_1, and similarly R_2, are not equal to R_r for the obvious reason that they have mismatched schemes. Even if they had matching schemes (just the attribute Name), they are most likely to have different sets of rows, and thus should not be equal. Technically, for the two tables R_1 and R_r, if the conjunct

```
SELECT 'true'
WHERE
        NOT EXISTS          AND    NOT EXISTS
             (SELECT *                  (SELECT *
             FROM R1                    FROM Rr
             EXCEPT                     EXCEPT
             SELECT *                   SELECT *
             FROM Rr)                   FROM R1)
```

holds true, then the two tables are identical in scheme and content. The problem is, in MySQL, EXCEPT requires the order and names of the columns to be identical which is complicated because students potentially could use arbitrary order and apply renaming of the columns. While MINUS does not impose these restrictions, it is not available in MySQL which is what we use as our back-end

[1] It is not just because the student did not include the PetID in the projection, but because the query as written will also include all the Purina brand FoodIDs in the inner projection and thereby making all the non priciest foods also eligible. The reference query R_r first finds the price of the highest priced Purina item, then picks the item's FoodID with an extra join before joining with *Likes* and *Pets* to avoid picking all the lower priced items unlike the student's queries R_1 and R_2.

storage, and thus does not work. While there are other computational[2] and query equivalence theory based [4] solutions, they usually require custom case by case programming, or impose significant restrictions on the class of queries, which is a major hurdle in designing a general good-for-all solution. For example, the ViSQL [11] SQL tutoring system based on Cosette [4] has significant limitations on the class of queries it can support, and was found to be handicapped as a first database class.

To avoid excessive custom coding not guaranteed to work well, we have designed an unorthodox walk around solution to equivalence checking problem that works even though it is slightly inefficient. Given that in a database class, the example databases used are small, and the queries often compute a very small sized table, the solution described below works perfectly well and flawlessly. We first concatenate all the rows in the two tables after converting them into strings, and create two giant strings. Then we sort the characters in the giant strings before generating a hash value using MD5 hash function. The expectation is that if the tables are identical, the hash values will be too. This approach is not impacted by column renaming or order.

4 Implementation of ReliQ

The IDE and the editor in Fig. 1 is the front-end of the ReliQ relational algebra query processor. Though we have abused the usage of the term ReliQ tutoring and assessment system, ReliQ is fundamentally a query processor. What makes it a tutoring and assessment system is its inclusion in the larger Project 360 engine. Project 360 is implemented using HTML, JavaScript, PHP, and MySQL to create a responsive, user-friendly, and feature-rich web-based platform. It consists of two components: the front end and the back end. The front-end of this platform is designed using the Bootstrap HTML framework, coupled with the versatile jQuery JavaScript framework.

The back-end system, on the other hand, is powered by the PHP Laravel framework, serving as the server-side scripting language. The foundation of this platform is the well trusted and widely used relational database management system MySQL. MySQL offers effective tools for data storage, retrieval, and modification, making it a prudent option for managing challenging database tasks. Also its compatibility with Laravel ensures smooth integration and seamless data interactions between the back-end and the database. We leverage Laravel's built-in authentication system to implement secure user registration, login, and password management.

The relational algebra subsystem ReliQ is implemented as a user friendly editor. The main part of the tool's body comprises a highly customized a WYSIWYG rich text editor, called CKEditor, specifically tailored to support relational algebra operators. The editor allows users to write and compose algebraic expressions directly within the web application. A third panel is dedicated to displaying the execution results as shown in Fig. 1 using PHP on the server-side.

[2] https://tinyurl.com/3vs8xkjr.

4.1 Syntax and Semantics

ReliQ supports all standard and extended set of relational algebra operators, and adopts the operator sets of Silberschatz, Korth and Sudarshan [19]. We have iconized the operators that can be clicked to select and place at the editing pane where the cursor is. All unary operators (e.g., select σ, project Π and rename ρ) uses two sets of parentheses – one set for the descriptors (Boolean condition, attribute list or the scheme) and the other for the operand, e.g., $\sigma_{()}()$. Users are able to fill those appropriately. Relation and attribute names can be selected by clicking on the table name or attribute names in the right, The subscripting of the descriptors are optional, and can be toggled by clicking on the X_2 icon at the far right end in the icon panel. Logical operators can also be selected or typed. In the current edition, the parentheses as described are mandatory.

The parentheses requirement is also true for binary operators such as union, intersection and join. For example, *Foods* $\cup$ *Likes* is syntactically incorrect. The correct syntax is (*Foods*) $\cup$ (*Likes*). This is because we are applying uniform bracketing rules across an expression. Since a binary operator can also be flanked by a pair of complicated expressions, we require that they be enclosed within a pair of parentheses to avoid confusion. Thus, the same rule applies to base tables as well. We are considering relaxing the parentheses rules in the future edition of ReliQ as it could become unwieldy at times. Yet, we believe that ReliQ syntax is more intuitive, forgiving, and easy to use than RelaX, radb, Relation or IRA. Furthermore, the group by ($\mathcal{G}$), assignment ($\leftarrow$) and order by (τ) have a more user friendly syntax in ReliQ than any of its predecessors.

4.2 ReliQ Query Processor

As discussed in Sect. 3.2, ReliQ query processor is implemented in MySQL using query translation. We did not include all the mapping algorithms from relational algebra queries to SQL in this paper for the sake of brevity. However, implementation of the assignment operator warrants special consideration and a brief discussion.

The MySQL back-end used by ReliQ for database support and SQL engine can only use one instance of the database. That essentially means all the users logging in to use a specific database are using one single instance. For querying purposes it is not a problem most of the time. However, the assignment operator is implemented using the CREATE VIEW statement at the SQL level. Also, CREATE VIEW in MySQL makes the generated view permanent for everyone. This situation has two implications. First, we cannot just drop a generated view for a given user to clean up the cache because another student may have generated an identically named view, which is truly common for assignment statements, e.g., $x \leftarrow exp$. The flip side of this issue is that once a user created a view named, say x, an assignment statement by another user trying to create another view named x will fail.

To resolve this issue, we generate a unique ID to the millisecond level and append to each relation name used to the left of the assignment operator. We also create and maintain a list of these tables, and drop the tables at the end of the execution of the query to clean up the cache. This strategy helps to manage the growth of useless tables in the back-end, avoid accidental dropping of active views, and conflicts with views generated by multiple users.

5 ReliQ as an Online Tutoring and Assessment System

ReliQ primarily supports two types of registered users – students and instructors. Students and instructors are allowed access to the public partition of ReliQ where instructors could post assignments, and tests over a test database. All students will have access to these artifacts for self-learning without any assistance from the instructor. However, instructors can open a course, and let students register in their courses in a more closed setting. A course can have defined duration, structure, help and more guided learning. In this section, we discuss the various components of ReliQ that support tutoring and assessment of relational algebra eLearning.

5.1 Instructor's Dashboard

Figure 3 shows the instructor's course dashboard through which she manages the course CS360 she created for her students. Aside from the fact that an instructor is able to do standard functions such as post syllabus, lecture notes, and learning materials, she also can post assignments, and tests, and grades. Finally, she is able to add students to her class. However, for the sake of brevity, we will only focus on how an instructor creates assignments and tests, and grades them.

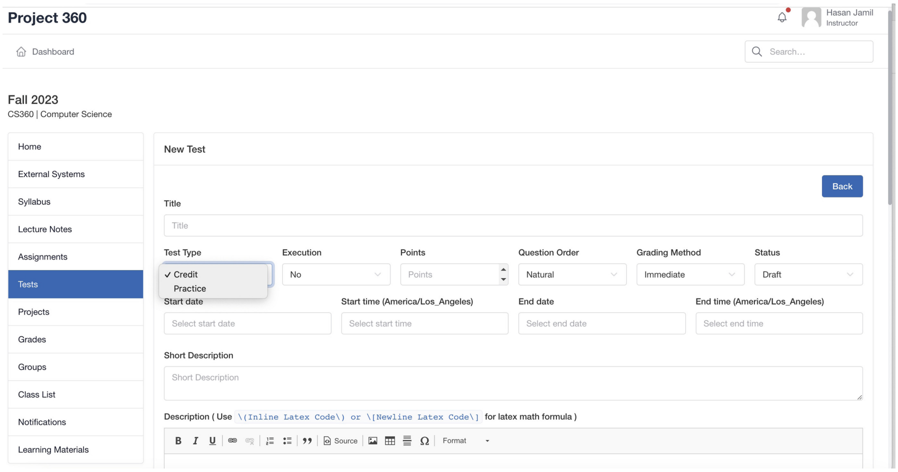


Fig. 3. Instructor's course dashboard.

Creating Assignments and Tests. One of the main features of ReliQ is designing and administering relational algebra assignments and tests, for both practice and credit. Figure 3 shows the preamble screen using which the instructor designs a test[3]. A test can be designed as a practice test or a for credit test by choosing the option from the drop down list on the left (e.g., *Credit* as shown). Test date and time, total points and the presentation order of questions can be selected. It can be saved as a draft, or published. Two additional features ReliQ supports uniquely are the options to allow query executions on a sample database, and a mode of grading. There are two modes of grading – immediate and deferred. In immediate mode, a grade is returned fully automatically at a preset time, usually after all students have finished taking the test. In deferred mode, however, the response to the tests are collected and saved, and not graded until the instructor manually requests an automated grading.

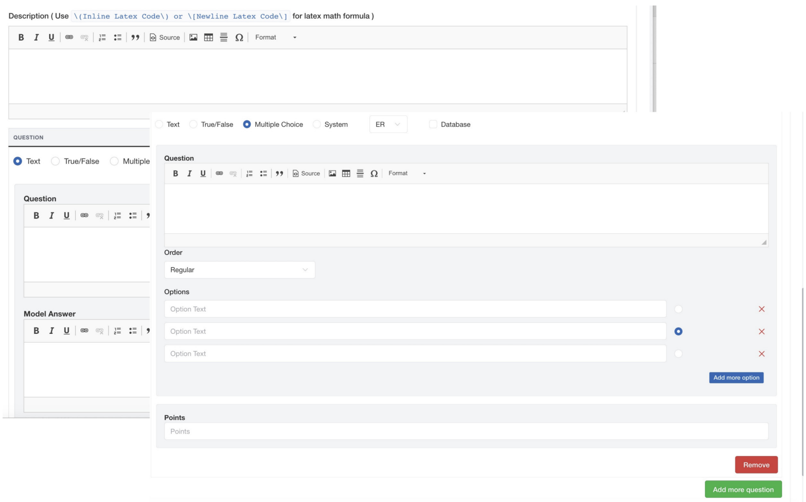


Fig. 4. Assignment creation – text response (inset) and MCQ response (foreground).

Similar to many learning management systems such as Canvas, ReliQ also supports creating questions in an interactive manner as shown in Fig. 4. As shown in Fig. 4, ReliQ supports plain text, "True/False" (not shown in Fig. 4) and MCQ type responses. However, it also supports Code or System type responses that can be directly executed in the back-end database systems. Every question is classified along three axes (if Code type response is selected) – subject category (ER, RA, SQL, FD), database inspection (Yes/No), and execution permission (Yes/No). For subject category "RA", database access selection "Yes" means that for this question, the student will be able to inspect the sample database tables to help her conceptualize query construction. Otherwise, she will have

[3] The process of designing an assignment is almost identical, and hence omitted.

to use her imagination to craft the query. On the other hand, if "Execution" is allowed, the student is allowed to both inspect the database tables and test run her constructed queries on the sample database since permission to execute queries is superseding to database inspection permission. But the converse is not true. Finally, the execution permission at the individual question level is locally overriding permission of the global execution choice the instructor makes in the preamble.

Depending on the response type, the panel for model or reference responses change. Figures 4 and 5 show the model response panels respectively for Text and MCQ, and RA System type responses. The ReliQ IDE shown in Fig. 1 appearing as a panel allows the construction of a relational algebra query, test run it against the sample database, and its correctness confirmed[4]. The model responses are used to auto grade tests and assignments and are never showed to the students unless instructor granted the permission.

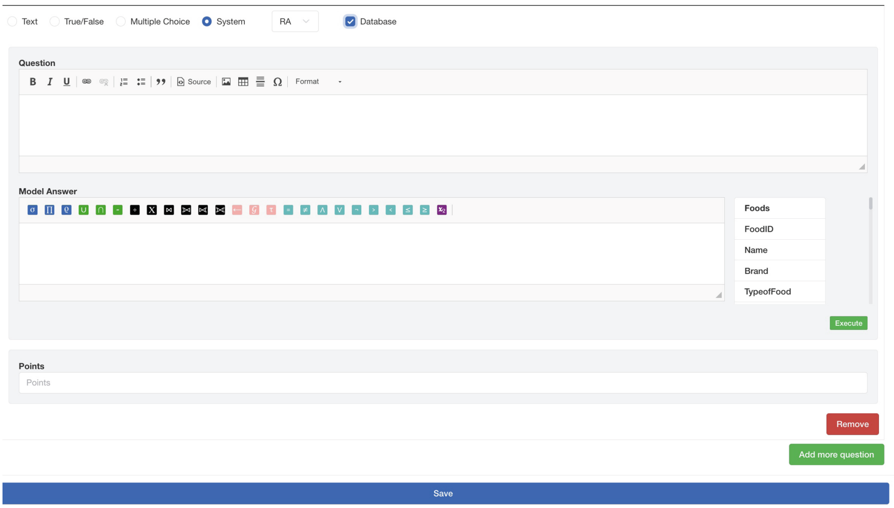


Fig. 5. RA System model response interface.

Automatic Grading. To be able to auto grade an assignment or test, all questions must be of any type except "Text" type responses, requiring manual assessment, in either immediate or deferred grading modes. While the True/False and MCQ type responses are easily graded using standard methods, the RA System questions are evaluated using the approach described in Sec 3.2 against the model response supplied by the instructor.

[4] Note that the ReliQ IDE also serves as a standalone interface for query construction and execution.

5.2 Student's Dashboard

Student's dashboard allows the students to browse published courses, register for courses, and attend the classes asynchronously using ReliQ (as part of Project 360). Additional functions allowed in the course interface are *Syllabus, Projects, Groups, People* and *Notifications*. Navigating to a registered courses, she can see course interface in Fig. 6 after selecting *Tests* (similarly for *Assignments*). The list displays brief descriptions of the test, and various other information related to the test, including test type and time frame when to take it. If already taken, it shows the points earned. It has two buttons – "Start" and "View". A test obviously cannot be viewed before the start of the test, but any time after the start time.

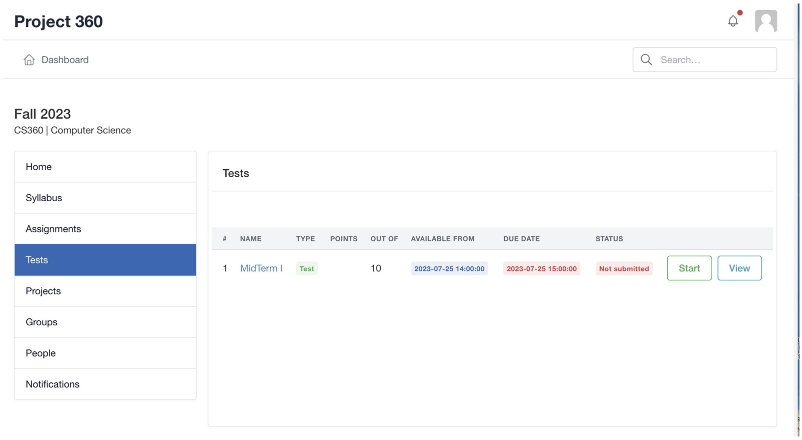


Fig. 6. Student's course dashboard.

The view of the tests (or assignments) students have is similar to the views of the Figs. 4 and 5. In such test interfaces, they are able to respond to test questions in ways similar to Canvas for Text, True/False and MCQ type questions. But distinctively, in Project 360, they are also able to write executable relational algebra queries directly in the response box for "System" or "Code" type questions, test run them, and save if allowed by the instructor.

6 Conclusion and Future Research

The main focus of this presentation was to highlight the design, functions and features of ReliQ tutoring and assessment system. We have largely emphasized how ReliQ supports construction and execution of relational algebra queries and tested for correctness. As an execution engine, it is similar to RelaX, but its inclusion into Project 360 afforded it more capabilities than any other known contemporary relational algebra tutoring system.

This cloud based ReliQ tutoring system is currently open to instructors for use in their database classes. We intend to collect usage and performance data to share with the community to improve database teaching and learning. We are distinctly aware of the US federal Family Educational Rights and Privacy Act (FERPA) law and keenly compliant with student data privacy. We therefore do not collect student personal identification data, though we collect demographic data to study social and learning related diversity, equity, inclusion, and accessibility (DEIA) issues. We therefore require students personal information to be mapped to non-discernible identities before submitted to the system by the instructors as course participants. Login credentials are then forwarded to the instructors for onward transmission to the students to gain access.

We have plans to enhance ReliQ with semantic feedback capabilities in the direction of RATest [15]. In the current edition, we show the standard SQL and MySQL versions of the relational algebra query. We have plans to also display the constructed query tree for better semantic comprehension of the structure of the queries. Finally, we are considering a feedback system similar to QueryVis [14] and viSQLizer [5] in the near future. These are some of the ideas we plan to seek as our future research.

Acknowledgement. This research was partially supported by an Institutional Development Award (IDeA) from the National Institute of General Medical Sciences of the National Institutes of Health under Grant #P20GM103408.

References

1. Brinkhuis, M.J.S., Savi, A.O., Hofman, A.D., Coomans, F., van der Maas, H.L.J., Maris, G.K.J.: Learning as it happens: a decade of analyzing and shaping a large-scale online learning system. J. Learn. Anal., **5**(2) (2018)
2. Brown, L.E., Feltz, A., Wallace, C.: Lab exercises for a discrete structures course: exploring logic and relational algebra with alloy. In: ITiCSE 2018, Larnaca, Cyprus, July 02–04, 2018, pp. 135–140. ACM (2018)
3. Chan, H.C.: Translational semantics for a conceptual level query language. J. Comput. Sci. Technol. **10**(2), 175–187 (1995)
4. Chu, S., Murphy, B., Roesch, J., Cheung, A., Suciu, D.: Axiomatic foundations and algorithms for deciding semantic equivalences of SQL queries. Proc. VLDB Endow. **11**(11), 1482–1495 (2018)
5. Folland, K.A.T.: viSQLizer: using visualization for learning SQL. In: NIK 2016, Høgskolen i Bergen, Bergen, Norway, November 28–30 (2016)
6. Gansner, E., Horgan, J.R., Kintala, C.M.R., Moore, D.J., Surko, P.: Semantics and correctness of a query language translation. In: DeMillo, R.A., editor, ACM POPL, Albuquerque, New Mexico, USA, January 1982, pp. 289–298 (1982)
7. Jamil, H.: Online tutoring and plagiarism-aware authentic assessment of database design assignments. In: IEEE TALE, Auckland, New Zealand, November 27-December 1 (2023). In press
8. Jamil, H., Shawon, F.: Automatic and authentic eassessment of online database design theory assignments. In: ICWL 2023, Sydney, Australia, November 26–28 (2023). In press

9. Ji, S., Yuan, T.: Conversational intelligent tutoring systems for online learning: what do students and tutors say? In: Kallel, I., Kammoun, H.M., Hsairi, L., editors, IEEE EDUCON 2022, Tunisia, March 28–31, 2022, pp. 292–298 (2022)
10. Jukic, N., Vrbsky, S., Nestorov, S., Sharma, A.: Erdplus. https://erdplus.com/ (2020). Accessed 31 July 2022
11. Karimzadeh, M., Jamil, H.: An intelligent online SQL tutoring system. In: IEEE ICALT 2022, Bucharest, Romania. July 1–4, 2022, pp. 212–213 (2022)
12. Kenny, C., Pahl, C.: Automated tutoring for a database skills training environment. In: Dann, W.P., Naps, T.L., Tymann, P.T., Baldwin, D., editors, ACM SIGCSE 2005, St. Louis, Missouri, USA, February 23–27, 2005, pp. 58–62 (2005)
13. Kessler, J., Tschuggnall, M., Specht, G.: RelaX: a webbased execution and learning tool for relational algebra. In: Grust, T., et al., editors, BTW 2019, 4–8 March 2019, Rostock, Germany, volume P-289 of LNI, pp. 503–506 (2019)
14. Leventidis, A., Zhang, J., Dunne, C., Gatterbauer, W., Jagadish, H.V., Riedewald, M.: QueryVis: logic-based diagrams help users understand complicated SQL queries faster. In: SIGMOD, pp. 2303–2318. ACM (2020)
15. Miao, Z., Roy, S., Yang, J.: Explaining wrong queries using small examples. In: Boncz, P.A., Manegold, S., Ailamaki, A., Deshpande, A., Kraska, T., editors, SIGMOD 2019, The Netherlands, June 30 - July 5, 2019, pp. 503–520 (2019)
16. Michel, F., Faron-Zucker, C., Montagnat, J.: A generic mapping-based query translation from SPARQL to various target database query languages. In: WEBIST 2016, Volume 2, Rome, Italy, April 23–25, 2016, pp. 147–158. SciTePress (2016)
17. Muehe, H.: Interactive relational algebra in latex. https://db.in.tum.de/people/sites/muehe/ira/ (2022). Accessed 22 July 2023
18. Murad, D.F., Wijanarko, B.D., Leandros, R., Murad, S.A.: The effectiveness of online-based authentic learning assessment. In: SIET 2022, Malang, Indonesia, November 22–23, 2022, pp. 277–282. ACM (2022)
19. Silberschatz, A., Korth, H.F., Sudarshan, S.: Database System Concepts, Seventh Edition. McGraw-Hill Book Company (2020)
20. Song, H., Kim, A., Park, S.: Translation of natural language query into keyword query using a RNN encoder-decoder. In: ACM SIGIR, Shinjuku, Tokyo, Japan, August 7–11, 2017, pp. 965–968. ACM (2017)
21. Tomaselli, S.: Relational - educational tool for relational algebra. https://github.com/ltworf/relational (2022). Accessed 22 July 2023
22. Wang, J., Stantic, B.: Facilitating learning by practice and examples: a tool for learning table normalization. In: BCI 2019, Sofia, Bulgaria, September 26–28, 2019, pp. 35:1–35:4. ACM (2019)
23. Yang, J.: IRA - interactive relationale algebra. https://users.cs.duke.edu/junyang/radb/ (2022). Accessed 22 July 2023
24. Zhang, M., Gu, Z.X., Amer, A., Sidhu, G., Srinivasan, S.: Software suite for self-paced learning. Int. J. Emerg. Technol. Learn. **17**(19), 20–32 (2022)
25. Zhekova, M., Pashev, G., Totkov, G.: An algorithm for translation of a natural language question into SQL query. In: ICERIS, Plovdiv, Bulgaria, October 13–14, 2022, volume 3372 of CEUR Workshop Proceedings, pp. 32–40 (2022)

Automatic and Authentic eAssessment of Online Database Design Theory Assignments

Hasan M. Jamil(✉) and Farjahan R. Shawon

University of Idaho, Moscow, ID 83844, USA
jamil@uidaho.edu, shaw0901@vandals.uidaho.edu

Abstract. One of the main reasons MCQ assessments are popular is that tests using text understanding is difficult and often erroneous. The emergence of large language models such as ChatGPT is not mature enough to help grading engineering assignments yet. Since MCQ tests are not well suited for summative assessment, scaling up eLearning for large number of students is difficult using non-MCQ tests. In this paper, we introduce a new eAssessment tool for database design courses that uses graphical conversations to understand learner's mental model of cognitive state. We show that the model is capable of substituting NLP for authentic assessment for eLearning.

Keywords: Intelligent tutoring · authentic assessment · eLearning · self-paced learning · user interface · graphical conversation · NLP substitute

1 Introduction

Recent research suggest that among various types of technologies, e-learning is the most commonly validated mode of delivery, followed by m-learning, Learning Management Systems (LMSs), and social media services [8]. While technology adoption, LMS adoption in particular [15], in higher education is increasing significantly in the post-Covid era, barriers still exist that prevent effective use of these tools. These barriers largely include instructors' lack of confidence, and preparedness, and less than enthusiastic acceptance of the available tools [4]. It is, therefore, not surprising that the instructors use these technologies mainly for blended learning [1] in which assessment[1] accounts only for 19% [17], perhaps because the instructors believe that the existing technologies are not a good fit for their style or personality [14], a factor that motivated the design of the system and research discussed in this article.

[1] According to Rhodes et al. [17], the main usage of LMSs can be broken down in different categories as follows: Announcements (82%), Items (77%), Grade (71%), Folders (62%), Files (53%), Assignments (53%), Web links (30%), Plagiarism detection (22%), Discussion boards (21%) and Tests (19%).

H. Xie et al. (Eds.): ICWL 2023, LNCS 14409, pp. 77–91, 2023.
https://doi.org/10.1007/978-981-99-8385-8_7

While there are tutoring systems for many CS subjects, there are significantly fewer systems for a first database course. The ones that are available, they split the topics into multiple tutoring systems (e.g., SQL [18], conceptual database design [11], relational algebra [13], and normalization [20]). We are, however, unaware of a single comprehensive tutoring system for database teaching. Using multiple independent systems to teach a single subject introduces substantial impedance mismatch, management hurdles and potential equity issues. To address these complex set of issues, we have initiated *Project 360* as a comprehensive database tutoring system. It is designed to include four inter-connected and integrated tutoring systems – Conceptual Database Design (*CoDD*), Visual SQL (*ViSQL* [12]), Relational Algebra Query Language (*ReliQ* [10]), and Normalization and Database Design Theory (*NoDD* [9]). In the remainder of this article, we discuss how NoDD approaches plagiarism averse authentic assessment and tutoring of functional dependency (FD) theory and database normalization.

2 Related Research

As opposed to a tutoring systems for teaching conceptual database modeling or database query languages, teaching functional dependency theory and database normalization probably are technically simpler. This is because these concepts are well defined and have established algorithms to compute them. As can be expected, there are several online normalization tools, or calculators, that do pretty well in computing various steps and components of dependency theories and normalization [19,20], and perform not too poorly in explaining the steps to the students. Unfortunately, that also means many existing online calculators, and more recently ChatGPT, make it extremely challenging to administer an online test without the risk of a dishonest student cheating [5].

From the tutoring view point, the concepts of functional dependency theory and normalization, the challenge is in designing a system that can get the point across, guide the students to learn the theories, and improve learning outcomes overall. There are several online systems that actually are effective at varying degrees. When it come to assessment, to the best of our knowledge, there is probably no online system that covers authentic assessment. Even if one existed, we believe that would be hugely vulnerable to plagiarism and cheating and will go undetected. *NoDD* most likely is the first attempt to rectify the shortcomings of contemporary online database normalization tutoring and assessment systems.

There have been several few attempts at building database normalization tools. However, only a handful of them have been designed for tutoring [20], and none are designed for assessment even though it is an integral component of learning. Essentially, assessment is left as an offline exercise. Furthermore, among the online normalization calculators, only the Normalization Tool [20], and Cho's normal form calculator [7] are live.

We also believe that an effective normalization theory learning tool must introduce to the students the fundamental concepts of functional dependencies, the inference rules and cognate theories so that they are able to derive closure of a set of a dependencies (F^*). This fundamental understanding serves as the foundation for them to follow the notions of attribute closures (X_F^+), which can be used to compute candidate keys, covers, and to test loss-less join decomposition.

While there are a few tools that could potentially help compute some of these, most are not comprehensive, and often only cover a subset of computing needs. Even when they do, are not in a form that is suitable for a learning system such as ours. For example, for the set of functional dependencies $F = \{A \rightarrow CD, B \rightarrow DE, AG \rightarrow BC, AB \rightarrow G, BG \rightarrow A\}$ over the scheme $R = (ABCDEGH)$ the FD calculator developed by Chakravarty [6] does not compute the normal forms. While it computes attribute and functional dependency (FD) closures, and minimal covers, it computes everything exhaustively and unnecessarily without offering any insight or explanation. An interesting FD calculator by Cho [7] also generates explanations of all its derivations. However, it does all the computations in one go in a non-interactive fashion. Unfortunately it is also not being maintained.

However, none of these systems are open sourced and no support is offered. Furthermore, they are designed for online direct interactions by the users and no APIs are supported. Thus considering all the aspects, the Normalization Tool [20] appears to be a standard calculator because it has shown a higher commitment to maintenance even though it does not offer all the functionalities we desire. It can optionally show the derivation steps, but without any explanations. Similar to FD Calculator, this is also modular and interactive. As we will highlight in the upcoming sections, NoDD also covers everything Normalization Tool does, and much more, but in a significantly unique way.

3 Database Design Assignments

Before we discuss the graphical conversation technique we have developed for eAssessment, it is perhaps helpful to discuss the nature and uniqueness of database design assignments, the subject of our eAssessment tool. One of the steps in database design is called schema normalization. A database scheme is a list of attributes, or column names of a table, and a set of functional dependencies that apply. To illustrate the concept of normalization, let us consider Example 1 below.

Example 1. R is a relation scheme over attributes $\{A, B, C, D, E\}$, and associated functional dependencies $F = \{AB \rightarrow C, A \rightarrow C, C \rightarrow A, CD \rightarrow A\}$.

R is said to be non-3NF compliant, and a decomposition into a 3NF scheme is required. This is done by discovering the canonical cover of the dependency set F, and following the 3NF decomposition algorithm and yielding two schemes – $R_1(AC)$, and $R_2(ABDE)$.

The tools instrumental in this decomposition are concepts of Armstrong's Axioms and Inference Rules, candidate key identification, attribute closure X_F^+, canonical cover, loss-less join decomposition and 3NF decomposition algorithm.

a 3NF decomposition of a scheme ensures that all functional dependencies are preserved and the splitting of the scheme is also loss-less. The decomposition of R into R_1 and R_2 above was accomplished first by determining that the dependency $CD \rightarrow A$ is redundant because $F \setminus \{CD \rightarrow A\} \models CD \rightarrow A$, i.e., even when $CD \rightarrow A$ is removed from F and $A \in CD^+_{F\setminus\{CD\rightarrow A\}}$. Then recognizing that $AB \rightarrow C$ is left-redundant. These two observations reduced F to the set $F_c = \{A \rightarrow C, C \rightarrow A$, and is called the canonical cover of F. We can also use standard techniques to discover the set of all candidate keys from either F or F_c easily. The complete set of candidate keys K of R is thus $ABDE$ and $BCDE$. Based on F_c and K, using the 3NF decomposition algorithm we obtain R_1 and R_2.

4 Graphical Conversational Interface

Our goal in the design of our graphical conversational interface for eAssessment is to construct a set of graphical alphabets or words with predefined meanings from which students will choose to write sentences with place holders for values. We then use a grammatical parsing system to understand the sentences to determine admissibility – or correctness. These sentences then can be assembled into a description of a solution. It turns out that the vocabulary can be compartmentalized based on the algorithms described in Sect. 5.

4.1 Language for Graphical Conversation

ChatBots, smart interfaces and many question answering systems use various forms of natural language conversations to engage with the users. While contemporary generative AI ChatBots actually construct responses live with a reasonable understanding of user questions, most other conversational systems are simpler and responses are often predetermined. For example, the retired Expedia travel help ChatBot[2] was largely a keyword and short phrase comprehension based decision tree workflow execution engine that was capable of performing simple routine tasks such as cancelling a reservation, re-booking a travel, reimbursements inquiry, etc. Amazon's essentially uses a fixed dialogue system with deterministic steps and outcomes. The graphical, or visual conversation system we are considering is somewhat similar to the Expedia's keyword based ChatBot with markedly distinct features in which users make numerous decisions to drive the conversation, graphically, withing a defined search space.

In NoDD, every conversation is based on a specific task, and the task defines the alphabet and the language for the conversation, all using graphical artifacts such as radio buttons, check boxes, selections, etc. As the Fig. 1 shows, once the choice for Proof by Inference Rules is selected, the vocabulary and the conversational grammar is preset. In this case, the four FDs over the scheme R are the context, and the conversation is about derivation of inferences using the choices

[2] Expedia recently introduced ChatGPT based travel assistant.

of rule application at the top (radio buttons). The sentences are the FDs, and the grammar for the construction of sentences are defined by these inference rules. NoDDs job is to ascertain if the user is constructing a valid sentence based on the grammar. It logs each step, so that feedback can be generated and the paragraph (the set of sentences) written can be graded. The editor in Fig. 1 simply allows the users construct fixed formatted sentences using check boxes without offering too much help. Only help it is offering is that it is constraining attribute choices within the scheme R[3].

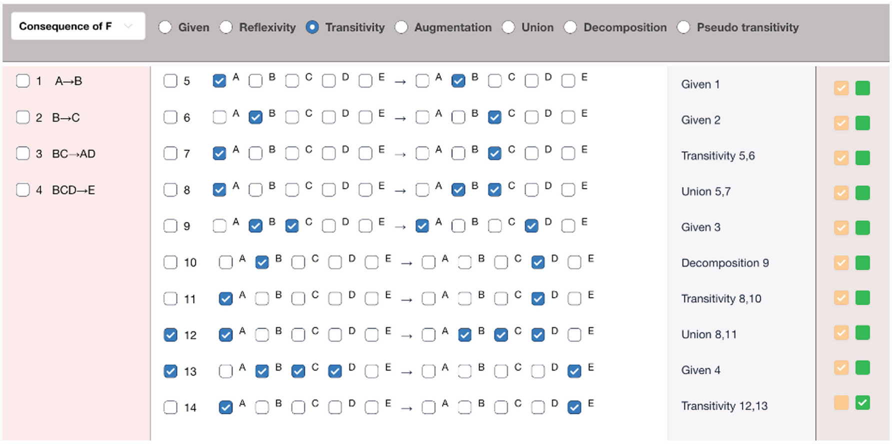


Fig. 1. NoDD graphical conversation for logical consequence of F.

4.2 NoDD Interface

The Normalization and Database Design tool we have designed is a unique and novel tutoring and authentic assessment system built for the sole purpose of database design theory teaching and learning. While there are several online calculator to help students solve database design theory related problems, none are comprehensive, and none can be used as an assessment tool. Technically, they also cannot be used as a tutoring tool, because they are designed to compute solutions directly without helping the students to do it themselves or removing their conceptual blocks. In contrast, the generic solution interface of NoDD, shown in Fig. 1, can be used for all FD related problems with slight adaptation for both tutoring and assessment.

[3] Removing this indirect help can be achieved by actually listing no attributes on both sides of the FDs, and letting the users write. For now, we believed that such an imposition does not buy much in terms of learning gains. Depending on user studies, this stance may change in the future.

The interface in Fig. 1 is partitioned into four quadrants. The top horizontal quadrant runs from the left edge to the right. This quadrant displays two types of menus and options selection. Users choose to solve or practice one of eight problem types from a drop down list on the left top quadrant. In the middle, there are seven radio buttons to choose one of the six inference rules, or choose any of the FDs in F. To the right, there are two buttons – one to create a new solution step, and one to conclude the solution. The remaining three quadrants are placed side by side just below the top quadrant. The leftmost quadrant lists all the active FDs with a serial number and a checkbox next to them. The middle quadrant is a workspace where users create solution steps using graphical alphabets. The rightmost quadrant primarily lists historical information of the solution steps. The sections below detail how this interface is used to solve the eight problem types in dependency theory. For the sake of brevity, in this presentation, we will only discuss the use of NoDD in assessment mode. The tutoring and practice mode is functionally similar but has important differences in how feedback is generated.

5 Functional Dependency Theory Algorithms in NoDD

Functional dependency theories are in the heart of database normalization. A sound understanding of its various components is essential for students to master normalization intricacies. Fortunately, all the concepts involved are algorithmically definable. In this section, we utilize those algorithms in the design of the conversational interfaces they must use to master functional dependency concepts and to help students learn normalization theory using NoDD.

5.1 Logical Consequence of FDs

Logical consequences of a set of dependencies F can be tested in two principal ways – by derivation using the Armstrong's axioms and inference rules, and by attribute closure computation. In both of these approaches, our principal goal is to understand if the student is following the correct algorithm by reconstructing it. Figure 1 illustrates how the process works in NoDD in reference to Example 2.

Example 2. Let the relational scheme be $R = (ABCDE)$ and the set of dependencies be $F = \{A \rightarrow B, B \rightarrow C, BC \rightarrow AD, BCD \rightarrow E\}$. Show that $F \models A \rightarrow E$.

Derivation Method. The graphical conversation proceeds as follows: To begin, the student selects the set of dependencies F, the scheme R, and the method she will use, and NoDD immediately prepares the interaction process as shown in Fig. 1 for $F \models f$ using derivation. NoDD displays the four dependencies of F and assigns the FDs a chronological number 1 through 4. Since a derivation method is chosen, a new FD is expected, and a new blank FD with a number 5 is displayed in the middle quadrant in the format shown below:

$\square\ f_5 : \boxtimes^{A}\square^{B}\square^{C}\square^{D}\square^{E} \rightarrow \square^{A}\boxtimes^{B}\square^{C}\square^{D}\square^{E}$ – Given 1 $\square$ $\square$

The derivation steps for FDs 6 and 7 are as follows:

$\square\ f_5 : \square^{A}\boxtimes^{B}\square^{C}\square^{D}\square^{E} \rightarrow \square^{A}\square^{B}\boxtimes^{C}\square^{D}\square^{E}$ – Given 2 $\boxtimes$ $\square$
$\square\ f_7 : \boxtimes^{A}\square^{B}\square^{C}\square^{D}\square^{E} \rightarrow \square^{A}\square^{B}\boxtimes^{C}\square^{D}\square^{E}$ – Transitivity 5 and 6 $\square$ $\square$

The result of the student choosing "Transitivity" in the third derivation step from the top panel indicating she is applying transitivity between FDs 5 and 6 is shown as f_7 above. The right quadrant comment shows what rules are being applied between the FDs 5 and 6.

However, the derivation is not yet complete since the student is yet to construct the derived dependency after applying transitivity between FDs 5 and 6, which should be $A \rightarrow C$. Student uses the check boxes on both sides of the dependency template f_7 to pick the attributes of R she believes should be in this new dependency. Figure 1 also shows the proper and complete derivation of the proof for $F \models A \rightarrow E$. Notice that, at every step until f_{14}, the yellow checkbox was selected to indicate that the derivation is incomplete at this step and a new derivation step is needed for the next derivation. Finally, the green checkbox was chosen at f_{14} to indicate completion of the proof.

Attribute Closure Method. When the "Attribute Closure" method is chosen, an interface simulating the derivation process below is presented. It starts by instantiating a derivation template as shown below:

$\{A\}^{+}_{F} = \{\square^{A}\square^{B}\square^{C}\square^{D}\square^{E}\}$ – FD $\square$ $\square$

Here too, the derivation follows systematically until the green checkbox is selected. The complete derivation and proof by attribute closure is shown below. In this process, students choose one dependency at each step that she believes is applicable, and adds or deletes the set of attributes believed to be in the closure.

$\{A\}^{+}_{F} = \{\boxtimes^{A}\square^{B}\square^{C}\square^{D}\square^{E}\}$ – Reflexivity $\boxtimes$ $\square$
$\{A\}^{+}_{F} = \{\boxtimes^{A}\boxtimes^{B}\square^{C}\square^{D}\square^{E}\}$ – FD 1 $\boxtimes$ $\square$
$\{A\}^{+}_{F} = \{\boxtimes^{A}\boxtimes^{B}\boxtimes^{C}\square^{D}\square^{E}\}$ – FD 2 $\boxtimes$ $\square$
$\{A\}^{+}_{F} = \{\boxtimes^{A}\boxtimes^{B}\boxtimes^{C}\boxtimes^{D}\square^{E}\}$ – FD 3 $\boxtimes$ $\square$
$\{A\}^{+}_{F} = \{\boxtimes^{A}\boxtimes^{B}\boxtimes^{C}\boxtimes^{D}\boxtimes^{E}\}$ – FD 4 $\square$ $\boxtimes$

Grading Student Responses. In both methods, the selections, rule applications, and the derivations are systematically logged as shown in the fourth quadrant history log. It is a simple process to determine if the student is incorrect, and all NoDD has to do is check if the final step is reached correctly. Although we are not discussing NoDD's tutoring mode in this article, it is not too hard to see how NoDD could offer feedback to the student in the event she made a mistake, which NoDD is able to detect at each step. However, designing effective hints is a research topic by itself, and will be presented separately.

5.2 Candidate Keys

One of the most important steps in database design is the discovery of candidate keys. However, to be able to determine if a relation scheme is in Third Normal Form (3NF) or Boyce-Codd Normal Form (BCNF), it is necessary to discover all candidate keys, not just one. Therefore, mastering a systematic derivation technique is essential. It turns out that candidate keys can be derived in three principal ways – exhaustive or iterative method, heuristic method and FD based reduction method. We present the tools we have developed for each below.

Exhaustive Method. In this method, a student basically tries every possible combination of the attributes in the scheme, and attempts to determine if the attribute closure recovers the entire scheme to declare the combination a superkey or a candidate key. The process is similar to the technique outlined in Sect. 5.1 for attribute closure computation. NoDD prepares a derivation template of attribute closure for all possible combinations, along with four check boxes – yellow, green, cyan and red, as shown below. As usual, yellow requests another step in the derivation process. Checking green indicates the current attribute set is a candidate key, cyan declares it a superkey, and red concludes it to be a non key, and the derivation for this combination stops.

$\{A\}^+_F = \{\Box^A \Box^B \Box^C \Box^D \Box^E\}$ – FD 1 ☐ ☐ ☐ ☐

This method is quite laborious and demands significant attention. A better method is the heuristic method described next that eliminates a large number of combinations that need to be tried, and thus shortens the process.

Heuristic Method. A sound heuristic that can be used to determine candidate keys involves separating the attributes into three distinct sets called the K^+, K^- and $K^?$. The attributes in a scheme that appear always on the left hand side in all FDs, or never appear in any FD, are placed in K^+ because they will always be in any candidate key. Similarly, the attributes that always appear on the right hand side, cannot appear in any candidate key since they are determined by other attributes, and thus belong to the K^- set. The remaining attributes are placed in the $K^?$ set because no decisions can be made about their status. The idea is to try attribute closures of all the combinations of the sets of attributes in $K^+ \cup K^?$, and not include attributes in K^- at all.

The derivation step follows the process in Sect. 5.2 preceded by the identification of these three sets, and displaying only the combinations involving the sets identified in K^+ and $K^?$. In Example 2, only E is always on the right hand side, and the remaining attributes are on both sides. Therefore, the identification follows the steps below:

$K^+ = \{\Box^A \Box^B \Box^C \Box^D \Box^E\}$ – FD ☑

$K^- = \{\Box^A \Box^B \Box^C \Box^D ☑^E\}$ – FD 4 ☑

$K^? = \{☑^A ☑^B ☑^C ☑^D \Box^E\}$ – FD 1, 2, 3, 4 ☑

NoDD therefore will list all combinations of $ABCD$ from the lowest cardinality combinations to the highest, and allow the process of key identification as in Sect. 5.2.

Reduction by Elimination Method. The elimination method though simple in concept, it is a complicated procedure to implement on a computer screen. It is basically a tree building process with the entire scheme of the relation at the root. Since all schemes are at least a superkey by itself, it can be used as a starting point for the reduction process. The idea is to use an FD in F to remove the attributes (say X) at the right hand side of it from the scheme at the root. This is because, X can still be recovered if the attribute closure of $R \setminus \{X\}_F^+$ is computed, and yet the scheme, or the superkey is reduced to a smaller superkey. The goal is to find the smallest superkey, the candidate key, by repeatedly applying this technique until no more attributes can be removed.

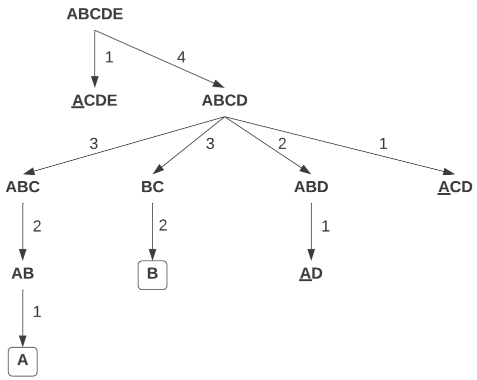


Fig. 2. Partial derivation of candidate keys using elimination method.

Figure 2 shows a partial expansion of the scheme R over the FDs F. The numbers on the arrows show the FD application. For example, using FD 1, the attribute B was removed from the root to create the node $ACDE$. The elimination process in this branch stopped because no other FD's have attributes in the left hand side that are a subset of $ACDE$ to help us eliminate them. Though it is a terminal node, it is not minimal as we see next. At the left tallest branch, we find a terminal node A (marked inside a box) which is minimal and was reached via repeated FD application in a similar way. Another boxed and terminal node is B. The other terminal nodes now can be checked to see if they are super sets of a boxed or any other terminal nodes and thus cannot be considered a candidate key. In this partial tree, all terminal non-boxed nodes are super sets of A.

This complex tree expansion and attribute elimination process is modeled in NoDD in a level wise fashion within the interface we have used so far. To initiate the elimination process, NoDD presents the root node as

1 : {ABCDE} – FD □ □ □ □

On clicking on the yellow checkbox, it generates the following display, where in the left is the root node, and at the right is a child of the root node at level 2.

1 : {ABCDE} – FD □ □ □ □ 2 : $\{\square^A\square^B\square^C\square^D\square^E\}$ – FD □ □

To deduce node $ACDE$ as shown in Fig. 2, the user chooses FD 1 at the left, checks B on the right node number 2. She also checks if this is a candidate key (green box), or is a superkey (cyan box). In this case, the user chose the cyan box, when the display reduces to the presentation below.

1 : {ABCDE} – FD □ □ ☑ □
2 : {ACDE} – FD 1 □ □ ☑ □

When the user selects the yellow box again on node 1, the numbering of the nodes changes and the following node structure is presented.

1 : {ABCDE} – FD □ □ ☑ □
2.1 : {ACDE} – FD 1 □ □ ☑ □
2.2 : $\{\square^A\square^B\square^C\square^D\square^E\}$ – FD □ □

On expansion in a similar manner, the structure below is generated.

1 : {ABCDE} – FD □ □ ☑ □
2.1 : {ACDE} – FD 1 □ □ ☑ □
2.2 : {ABCD} – FD 4 □ □ ☑ □

A complete expansion of the derivation tree will identify the two candidate keys A and B. The final node showing A as the candidate key will be presented as

2.2.1.1.1 : {A} – FD 1 □ ☑ □ □

5.3 Covers of Dependencies

Covers of dependencies require two sets of dependencies, say F and G. To establish the fact that F covers G, we are required to show that $\forall g(g \in G, F \models g)$. However, to show that $F \equiv G$, we must show that $\forall f, g((f \in F, G \models f) \wedge (g \in G, F \models g))$. For this task, we again repurposed the interface discussed in Sect. 5.1. This time, instead of one set of dependencies, we list two sets, and allow a checkbox to make one of them active, say F. The required step here is that the student must choose (the system does not require it) an FD $g = X \rightarrow Y \in G$ by checking the box next to it. Once she does, a dialog box appears as before in a slightly different manner as shown below:

$\{\square^A\square^B\square^C\square^D\square^E\}^+_F = \{\square^A\square^B\square^C\square^D\square^E\}$ – FD 1 □ □ □

This time, we require the user to choose the set of attributes (X) in the left of the equal sign for which the attribute closure is to be computed. Then follow the process in Sect. 5.1 with respect to the dependencies in F and demonstrate that she could recover Y in the right side of the equal sign. If she could not, she should check the red box and declare that F does not cover G. Instead, if $Y \subseteq X_F^+$, then she should choose to check the green box. Then repeat the same process for all the dependencies in G. Every time she checks the green box, the check box next to the FD in G also turns green, red otherwise. To show that $F \equiv G$, she must now switch F to G as the active FD set, and show that $G \models f$, for all $f \in F$.

5.4 Canonical Cover of FDs

Theoretically, the canonical cover of a set of FDs F is the set of FDs F_c such that there is no redundant FDs in F_c and no FDs has extraneous attributes in their left hand sides, i.e., minimal. To compute F_c, a student must follow the following three steps, in no particular order.

Decomposing FDs. While decomposition of FDs of the form $f = X \rightarrow Y$ into one attribute in the right hand side is not necessary, but doing so makes the remaining process simpler. However, combining or taking union of the smaller schemes into one using applicable FDs may become necessary to reverse the decomposition of FDs thus made, again not mandatory. This step can be achieved by choosing an FD, and clicking a button to decompose it. Once clicked, NoDD decomposes the FD into $X \rightarrow A$ form, and reorders and renumbers the FDs. The set of FDs generated to replace the FD $X \rightarrow Y$ is equal to the number of distinct attributes $A \in Y$. For example, the FD $BC \rightarrow AD$ is replaced by $BC \rightarrow A$ and $BC \rightarrow D$ using the simple application of the decomposition rule.

Redundant FD Removal. An FD $f = X \rightarrow A \in F$ is redundant, if F still implies f, i.e., $F \setminus \{f\} \models f$. That means, for an FD $f = X \rightarrow A$ to be non-redundant, $A \notin X^+_{F \setminus \{f\}}$ must hold. We again re-purpose the cover computation subsystem discussed in Sect. 5.3 with the following adjustments. To test if an FD is redundant, the student selects an FD f from the FD list F by checking the box next to it. She is then presented with the conversation below:

$$\{\square^A\square^B\square^C\square^D\square^E\}^+_F = \{\square^A\square^B\square^C\square^D\square^E\} \text{ – FD 1 } \square\ \square\ \square$$

and she must follow the procedure discussed in Sect. 5.3. At the end, she must check either “not redundant” (green), or “redundant” (red). Once checked, the button next to the FD in FD list is colored green or red, and the FD is unchecked. However, she can optionally check the FD checkbox again to remove it from the FD list, when it is greyed out. The process can be repeated for all the FDs in F.

Removing Left-Redundancies in FDs. To remove left redundancies, we use the same interface and for the redundant FD removal, but now turn on the left-redundancy removal radio button. Thus the multi-function red button is able to allow removal of attributes from the FD currently checked. This time, check boxes in the attributes in the FD appears, and the student is able to select the attributes she wish to remove. The checkbox next to the FD stays unchecked. This time too, if the FD checkbox is selected, the attributes checked in the FD is greyed out to indicate removal.

5.5 Normal Form Decomposition

During this step, students choose a set of dependencies F, and use the tools described in Sects. 5.2, and 5.4 to compute all the candidate keys K of R and the canonical cover F_c of F. While there are definitions for both 3NF and BCNF to test if a scheme is already in these normal forms, and by checking those conditions we could potentially save some time and avoid unnecessary decomposition, for the sake of brevity, we will not address this issue in this paper. In the following two sections, we present the process a student will follow to decompose a scheme into 3NF or BCNF assuming that the scheme warranted the decomposition[4].

Third Normal Form. Recall that the 3NF status of a scheme is related to transitive and partial dependence of attributes on non-key attributes. To avoid making mistakes by following a manual inspection, the algorithms for these two normal forms ensure a correct decomposition. The algorithm leverages the canonical cover computation and asks to create a distinct scheme S using the attributes in a FD $f \in F_c$ only if $S \not\subseteq R_i$ where R_i, $1 \leq i \leq n$, is one of the decomposed schemes already created, and then renaming S as R_{n+1}. Finally, if none of the R_is is a super set of any of the candidate keys in K, then create a final distinct scheme R_{n+1} using the attributes of one of the keys in K.

To respond to a test question on 3NF decomposition, students can start by computing the canonical cover and the candidate keys. Or, they can enter them in the system manually. Technically, the interface will have displayed a set of dependencies and a set of keys. Students will identify an FD or a key by checking the box, to work on it. NoDD presents the scheme as shown below that is tied to the FD or key selected. Selecting an FD displays

$$R_1 : \{\square^A \square^B \square^C \square^D \square^E\} - \text{FD } 1 \;\square\;\square$$

and the student selects the attributes for the decomposed scheme. On the other hand, selecting a key displays

$$R_1 : \{\square^A \square^B \square^C \square^D \square^E\} - \text{Key } 1 \;\square\;\square$$

[4] It must be noted here that even if the scheme is already in one of these normal forms, attempting to decompose it or doing so does not render the decomposition unsound, or introduce a design error. Finally, most often then not, the algorithms will return the original schemes anyway.

and the scheme is created as in the previous case. Once the green box is checked, the scheme is added as one of the decomposed schemes. Checking red box removes it from the set. We should note that we prefer not to aid in constructing the schemes by pre-selecting the attributes based on the FD, a key principle we have discussed as our design feature. This will defeat the purpose of testing the cognitive process of the students. We offer them enough so that they are able to pick the right statements, not responses.

Boyce-Codd Normal Form. BCNF decomposition if a complicated and involved algorithm. It is fundamentally a binary tree expansion algorithm. We, therefore, adapt an interaction system similar to candidate key discovery algorithm using reduction by elimination method discussed in Sect. 5.2. This time, we expect a right skewed binary tree, though we support a general tree expansion as before. Therefore, we simplify the numbering scheme for node identification with the following changes. The root node is displayed as follows along with its first possible child:

1 :	$\{ABCDE\}$	☐	☐
a :	$AB \rightarrow C$	☐	
b :	$A \rightarrow C$	☐	
c :	$CD \rightarrow A$	☐	
K_1 :	ABDE	☐	
K_1 :	ACDE	☐	

2.1 :	$\{\square^A\square^B\square^C\square^D\square^E\}$	– FD	☐
a :	$\square^A\square^B\square^C\square^D\square^E \rightarrow \square^A\square^B\square^C\square^D\square^E$	☐	
K_1 :	$\square^A\square^B\square^C\square^D\square^E$	☐	

Students are expected to choose one of the dependencies in the root node, and derive a child node based on the dependency just chosen. In this child node, the student also is expected to identify the FDs and keys that would hold. She uses the yellow check boxes to add one more dependency or key, and the green box to confirm the node and exit. By checking the attribute boxes, she constructs the scheme, FDs and keys.

Finally, once the child green box is checked, a new child node is created along with the information just supplied. Student can click on the parent's yellow box again to generate one more template for a new child node, and new description begins. This time, the child is assigned a node number 2.2, since there is already a number 2.1. Until the green box of the root or parent is checked, more children can be generated. Furthermore, when node 2.1 (say) is used as the root node, the new child is numbered 3.1 (child of a level 2 node), and all subsequent level three nodes are sequentially numbered.

6 Conclusion and Future Research

We believe there are two principal issues in authentic eAssessment we need to tackle. First, we need to use a language that is easily and correctly interpretable for the student to express her mental model of the problem being solved or the question she is responding to. Using a natural language as a modality of communication, such as English, is ideal in the age of large language models (LLM) and seems appropriate. However, as a recent conversation[5] with ChatGPT reveals,

[5] https://chat.openai.com/share/04aaf386-a76b-4ad7-b814-d840d139eb0b.

it is not that simple and it is far less perfect than one might expect. Leaving the costs of developing tailored technologies using LLMs as backbones aside, its efficacy and efficiency of human-AI communication for such purposes require further improvement.

Excellent alternatives to natural language based human computer communication are visual languages [2], and visual interfaces [16]. Visual interfaces are finite and simpler than visual languages, while visual languages are more open and difficult to design, they limit the scope. In this paper, we blended these two technologies into a language to capture the solution process a student follows in answering database design questions. Though it is limited to our application, it is highly focused and effective. In particular, it is easy to follow the chain of thought and easy to interpret the thought process for the purpose of grading.

Second, the approach we have adopted helps us stay plagiarism averse. Given the nature of the subject and the availability of a large number of online tools that can instantly provide the answers to all test questions, we needed to focus on capturing the analytical process followed by the test takers that they cannot glean from the internet calculators effortlessly. In the absence of a proctor in online test taking and the real possibility of cheating, forcing the student to display the depth of subject matter understanding and comprehending the explanations are critical issues. NoDD achieves that goal by offering a subject-grounded conversational language for question answering and interpreting the responses flawlessly.

In this paper, we refrained from discussing the algorithms for interpreting the responses for the assessment and feedback generation during tutoring sessions [3]. Though it is easy to sense that the NoDD interface diligently collects the sequence of thought process of a student, interpreting them requires careful attention to details. Feedback generation in the form of prompts and suggestions just enough to put the student on track also requires careful design and sophisticated models. These are some of the issues we seek to explore in our future research.

Acknowledgement. This research was partially supported by an Institutional Development Award (IDeA) from the National Institute of General Medical Sciences of the National Institutes of Health under Grant #P20GM103408.

References

1. Anthony, B., Jr., et al.: Blended learning adoption and implementation in higher education: a theoretical and systematic review. TKL **27**, 531–578 (2022)
2. Ayala, P., Barandiaran, I., Vicente, D., Graña, M.: Exploring simple visual languages for real time human-computer interaction. In: IEEE VECIMS 2003, Lugano, Switzerland, 27–29 July 2003, pp. 107–112 (2003)
3. Ball, M.A., Garcia, D.D.: Autograding and feedback for snap!: a visual programming language (abstract only). In: ACM SIGCSE 2016, Memphis, TN, USA, 02–05 March 2016, p. 692 (2016)

4. Buabeng-Andoh, C.: Factors influencing teachers' adoption and integration of information and communication technology into teaching: a review of the literature. IJEDICT **8**, 136–155 (2012)
5. Carr, N., Jamil, H., Shawon, F.: An experiment on leveraging chatGPT for online teaching and assessment of database students. In IEEE TALE 2023, Auckland, New Zealand, 27 November–1 December 2023 (2023)
6. Chakravarty, A.: Functional dependencies checker (2018). https://tinyurl.com/yrzrhcm3. Accessed 8 Jan 2022
7. Cho, R.: Tool for database design (2017). https://tinyurl.com/45jz7bpv. Accessed 8 Jan 2022
8. Granic, A.: Educational technology adoption: a systematic review. Educ. Inf. Technol. **27**, 9725–9744 (2022)
9. Jamil, H.: Online tutoring and plagiarism-aware authentic assessment of database design assignments. In: IEEE TALE, Auckland, New Zealand, 27 November–1 December 2023 (2023)
10. Jamil, H., Naha, K., Shawon, F.: An online tutoring and assessment system for teaching relational algebra in database classes. In: ICWL 2023, Sydney, Australia, 26–28 November 2023 (2023)
11. Jukic, N., Vrbsky, S., Nestorov, S., Sharma, A.: Erdplus (2020). https://erdplus.com/. Accessed 31 July 2022
12. Karimzadeh, M., Jamil, H.: An intelligent online SQL tutoring system. In: IEEE ICALT 2022, Bucharest, Romania, 1–4 July 2022, pp. 212–213 (2022)
13. Kessler, J., Tschuggnall, M., Specht, G.: Relax: a webbased execution and learning tool for relational algebra. In: BTW 2019, Rostock, Germany, LNI, 4–8 März 2019, vol. P-289, pp. 503–506 (2019)
14. Kimmerl, J.: Does who am I as a teacher matter? exploring determinants of teachers' learning management system adoption in education style. In ICIS 2020, Hyderabad, India, 13–16 December 2020. Association for Information Systems (2020)
15. Ngafeeson, M.N., Gautam, Y.: Learning management system adoption: a theory of planned behavior approach. IJWBLT **16**(1), 27–42 (2021)
16. Ohira, M., Masaki, H., Matsumoto, K.: CICRO: an interactive visual interface for crowd communication online. In: Ozok, A.A., Zaphiris, P. (eds.) OCSC 2011. LNCS, vol. 6778, pp. 251–260. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-21796-8_27
17. Rhode, J., Richter, S., Gowen, P., Miller, T., Wills, C.: Understanding faculty use of the learning management system. Online Learn. J. **21**, 68–86 (2017)
18. Röhm, U., Brent, L., Dawborn, T., Jeffries, B.: SQL for data scientists: designing SQL tutorials for scalable online teaching. VLDB **13**(12), 2989–2992 (2020)
19. Stefanidis, C., Koloniari, G.: An interactive tool for teaching and learning database normalization. In: Proceedings of the 20th Pan-Hellenic Conference on Informatics, Patras, Greece, 10–12 November 2016, p. 18. ACM (2016)
20. Wang, J., Stantic, B.: Facilitating learning by practice and examples: a tool for learning table normalization. In: BCI 2019, Sofia, Bulgaria, 26–28 September 2019, pp. 35:1–35:4. ACM (2019)

From Classroom to Metaverse: A Study on Gamified Constructivist Teaching in Higher Education

Peter H. F. Ng[1,2(✉)], Peter Q. Chen[1], Zackary P. T. Sin[1], Ye Jia[1], Richard Chen Li[1,3], George Baciu[1], Jiannong Cao[1], and Qing Li[1]

[1] Department of Computing, The Hong Kong Polytechnic University, Hong Kong, China
peter.nhf@polyu.edu.hk
[2] Department of Rehabilitation Sciences, The Hong Kong Polytechnic University, Hong Kong, China
[3] Department of Applied Social Sciences, The Hong Kong Polytechnic University, Hong Kong, China

Abstract. In the rapidly evolving educational landscape, the integration of metaverse and gamification is emerging as a revolutionary approach. This paper presents the Gamified Constructivist Teaching in the Metaverse (GCTM) framework, aiming to enhance engagement and satisfaction in the computer science education domain. Implemented in two engineering classes using the metaverse platform, the GCTM model, with its unique combination of game world design, rule design, roleplay, mission assignments, and evaluation, demonstrated promising results in enhancing student-lecturer interactions. Feedback indicated a stronger sense of belonging among students in the virtual environment compared to conventional platforms like ZOOM or MS Teams. The findings underscore the potential of GCTM in transforming the educational experience, suggesting a significant stride towards a more interactive and learner-centric approach in the metaverse-driven educational era.

Keywords: Metaverse · Eduverse · Constructivism · Online teaching

1 Introduction

The COVID-19 pandemic has precipitated unforeseen and sustained consequences on the system of higher education [1]. Over the course of several recent semesters, a pervasive shift has occurred from traditional face-to-face (F2F) instructional delivery to a synchronous online mode. This transformation has catalyzed the development and utilization of an abundance of multimedia resources, including numerous videos and online materials tailored for autonomous student learning [2].

The exigencies of the pandemic have compelled students to adapt to new paradigms of learning, which predominantly manifest in online or hybrid formats, as physical attendance became a logistical impossibility in many instances. Within our institutional

H. Xie et al. (Eds.): ICWL 2023, LNCS 14409, pp. 92–106, 2023.
https://doi.org/10.1007/978-981-99-8385-8_8

context in 2021, we implemented a synchronous hybrid model of instruction. This adaptation was met with varied levels of student engagement. For instance, when presented with the opportunity to attend classes in person, only a minority of 13.5% of the student body participated in all the face-to-face sessions.

A closer examination reveals a more disconcerting trend within specific academic units, such as the Department of Computing. As depicted in Table 1, only 5.6% of students in this department attended all the face-to-face classes. Furthermore, a subset comprising 4.9% of the students opted to engage with class recordings rather than participating in real-time instruction for more than half of the classes. This observation warrants further investigation into the underlying factors that may be influencing these patterns, potentially offering insights into the broader challenges and opportunities presented by this seismic shift in educational delivery.

Table 1. A Survey of Class Modality at Polyu in 2021

Class Modality	COMP N = 142	PolyU N = 3457
I attended all my classes f2f	5.6%	13.5%
I attended more than half of my classes f2f	18.3%	24.0%
I attended about half of my classes f2f	17.6%	16.8%
I attended more than half of my classes online	23.9%	25.1%
I attended all my classes online	28.9%	16.7%
I watched class recording instead of attending classes in real time for more than half of my classes	4.9%	2.4%

COMP = Students in Department of Computing
PolyU = Students in The Hong Kong Polytechnic University

In a subsequent survey conducted in 2022, the evolving preferences and behaviors of students with respect to instructional delivery were further elucidated. According to the data presented in Table 2, a mere 15% of the students surveyed expressed a preference for face-to-face class attendance. Within the specific disciplines surveyed, this preference manifested in 209 students in the Computer Science department (COMP) and 62 students in the Electrical and Information Engineering department (EIE).

This marked decline in preference for traditional classroom attendance reflects a potential paradigm shift in educational expectations and engagement. These findings invite further scholarly exploration into the underlying causes and potential implications of this trend, with particular attention to how institutional strategies and pedagogical practices must evolve in response to these changing dynamics. The implementation of online teaching modalities, such as ZOOM or Microsoft Teams, offers distinct advantages in terms of class management, encompassing functionalities like lesson recording, account administration, and security measures. While ZOOM is architected primarily for webinars [3], Microsoft Teams is oriented towards meetings [4]. Both platforms facilitate a form of passive learning, catering well to this mode of educational engagement.

Table 2. A Survey of Preferred Class Modality at Polyu in 2022

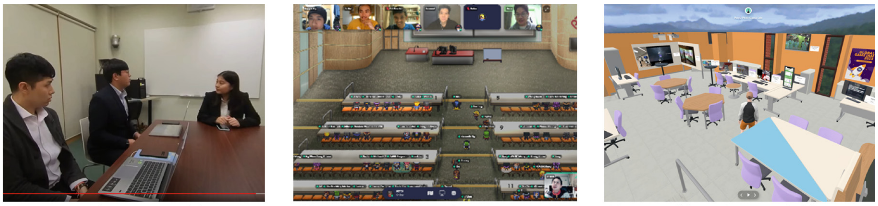
Class Modality	COMP N = 142	PolyU N = 3457
Entirely face-to-face	32(15.4%)	10(16.1%)
Minimal use of the Web, mostly held in face-to-face format	21(10.1%)	12(19.4%)
An equal mix of face-to-face and web content	68(32.7%)	19(30.6%)
Extensive use of the Web, but still some face-to-face class time	51(24.5%)	16(25.8%)
Entirely online with no face-to-face time	36(17.3%)	5(8.1%)

COMP = Year 3 students in Department of Computing
EIE = Year 4 students in Department of Electronic and Information Engineering

However, an intriguing and somewhat concerning observation was made in the transition to online education. It appears that the pandemic-induced shift to online learning has fostered a more passive approach among many students. Common behavioral indicators include reluctance to activate video cameras or engage in discussion with classmates, leading to a noticeable absence of peer-to-peer interaction and a diminished sense of community and belonging.

This trend is particularly disconcerting given the overarching philosophy that university education extends beyond mere knowledge transfer. Rather, the role of higher education is to cultivate active, lifelong learners capable of addressing real-world problems [5]. This mission inherently involves nurturing soft skills and fostering teamwork, capacities that can be inhibited if active and peer-to-peer learning is lacking [6, 7]. A deficit in these areas may impede the development of essential problem-solving abilities and undercut efforts to strengthen the core component of learn-to-learn skills.

In response to the aforementioned challenges, we initiated different innovative approach in 2020, leveraging mixed reality and the metaverse technologies to stimulate a more active learning stance among students during online instruction and various activities (see Fig. 1) [8–12]. We have endeavored to provide immersive and interactive experiences for the students. This paper elucidates three fundamental concepts underpinning our implementation. Firstly, it explores the methodologies employed to interconnect diverse technologies for the provision of education within the metaverse, a virtualized environment. Secondly, it examines strategies to engage students by employing gamification techniques and constructive teaching methods, thereby fostering collaboration, motivation, and active participation in the learning process. Thirdly, it investigates the integration of knowledge through the utilization of an interactive knowledge graph model, a sophisticated structure that links related information in a networked fashion.



(a) 360-degree video to understand the working situation in IT industry (Link)

(b) Teaching in Gather.Town, a 2D spatial video conferencing platform in a large class (N>200) (Link)

(c) Teaching in Spatial.io,a 3D spatial video conferencing platform in a small class (N<15) (Link)

Fig. 1. Applying different mixed reality and the metaverse technologies in teaching

2 Literature Review

2.1 Metaverse and Educational Metaverse

The concept of the Metaverse is subject to varying interpretations, reflecting its multifaceted nature. Jon Radoff has introduced an onion structure to represent the Metaverse, employing a seven-layer model to articulate the diverse requirements from infrastructure to user experience within this virtual realm [13]. Similarly, Brian Jackson of InfoTech Research Group has advanced a union structure for the Metaverse, positing it as a technological convergence [14]. Numerous analytics and market research firms have also contributed various infographics, delineating the key players and stakeholders within the Metaverse ecosystem [15]. Despite these attempts at definition, the complexity and multifarious nature of these representations may render them somewhat inaccessible for educators aiming to integrate metaverse technologies into their teaching practice. Further complicating matters, the broad scope and rapid evolution of the Metaverse leave us without a comprehensive understanding of how distinct technologies operate within education, why they are necessary, and how they may evolve in the near future [16].

2.2 Gamification and Constructive Teaching

Gamification refers to the application of game-related elements within non-gaming environments, seeking to harness the intrinsic motivators that games offer [17]. Contrasted with traditional learning approaches, such as lecture-based teaching, gamification presents several advantages that align with contemporary educational goals. These benefits include: 1) the enhancement of student motivation through interactive and competitive elements; 2) the potential for improved knowledge retention, as students often engage more deeply with gamified content; 3) the introduction of social mechanisms, such as badges, points, or leaderboards, which can foster a sense of community and achievement, thereby engaging students on a more profound level [18, 19]. The adoption of gamification within the metaverse opens new avenues for exploration, extending the boundaries of traditional learning and creating an enriched, dynamic educational landscape.

Constructivism is a pedagogical theory positing that learners actively construct knowledge rather than merely absorbing information in a passive manner. Through experiencing the world and reflecting on these experiences, individuals craft their own representations and integrate new information with their pre-existing knowledge structures [20, 21]. This constructivist approach has been employed in our service-learning projects, wherein students are immersed in real-world problems. The learning process involves understanding the problem, analyzing technical solutions, engaging with stakeholders, and devising solutions to authentic challenges [22]. We have observed that the constructivist approach is particularly resonant with engineering students and are thus motivated to explore its extension from the physical to the virtual world.

An instructive application of constructivist principles is embodied in the 5E Model, which was developed by the Biological Sciences Curriculum Study (BSCS) in 1987. The 5E Model is oriented toward promoting collaborative and active learning, with the goal of encouraging students to be engaged and social participants in their educational journey. By actively engaging with and reflecting on educational activities, students are facilitated in reconciling new knowledge with their existing conceptual frameworks. Several educational movements, such as inquiry-based learning, active learning, experiential learning, discovery learning, and knowledge building, are recognized as variations on the constructivist theme. According to subject matter expert Beverlee Jobrack, these methodologies are unified in their commitment to fostering an enriched, active engagement with learning content [23–25]. In synthesizing these principles with the emerging possibilities of the virtual world, the stage is set for a vibrant expansion of pedagogical innovation and effectiveness.

2.3 Connection with the Domain Knowledge

The knowledge graph is also considered a solution to reduce information overload during browsing domain knowledge for new knowledge users [26]. Hao et al. used a knowledge map and social network analysis to navigate the knowledge users. They used a Boolean value to connect each knowledge node, ignoring the similarity between domain knowledge. Janowicz et al. adapted a knowledge graph to visualise environmental knowledge and spatial data to a 2D knowledge graph [27]. Lin et al. try to turn medical expertise into a knowledge graph by representing knowledge with keywords or phrases [28]. Others might also try illustrating students' hobbies in phases or words and connect them with probabilities [29]. Those researches mainly focus on representing the domain knowledge with keywords or phrases and representing them on a 2D knowledge graph. However, with the development of VR devices and the metaverse concept, we could have another approach that visualises the domain knowledge in 3D space and integrate the knowledge in a novel way.

3 Methodology

3.1 Connect the Metaverse Service and Application

We posit that the future landscape will not be dominated by a singular Metaverse but rather will consist of an array of metaverse-related applications and developments. These will likely be interconnected through concepts such as digital identity and digital ownership. Given these considerations, we have sought to distill the concept of the Metaverse for educational purposes into a more accessible three-layer structure for educator (see Fig. 2).

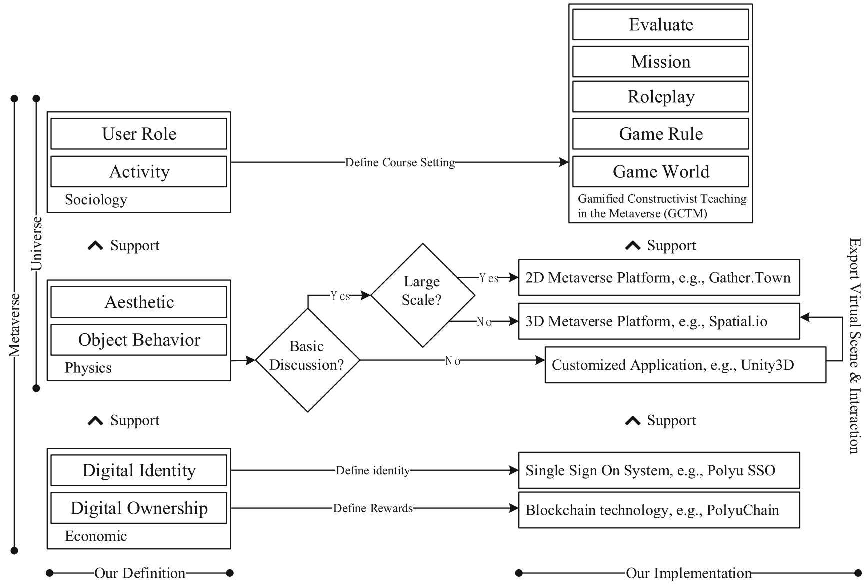


Fig. 2. Our Definition and implementation of Metaverse in Education

The uppermost layer in our conceptual framework for education within the Metaverse is the sociological layer, concerning the design of user roles and activities for students in the virtual world, essentially constituting the course setting. Upon entering the virtual world, students are represented through avatars, and thus, they anticipate adopting diverse roles to partake in various activities, akin to experiencing another reality. This layer emphasizes the importance of well-defined roles and structured activities, allowing students to engage meaningfully within the virtual space. We have articulated a specific framework to guide this complex design process, termed Gamified Constructivist Teaching in the Metaverse (GCTM), which will be elaborated upon in the subsequent section. By foregrounding the sociological aspects of interaction within the virtual realm, this layer facilitates an enhanced understanding of how students can immerse, engage, and learn through alternate personas and meticulously crafted activities, reflecting the multifaceted nature of human interaction and learning in virtual environments.

The intermediate layer of our conceptual framework pertains to the physical layer, i.e., virtual world, within the Metaverse, encompassing both the aesthetic and object behavior in the virtual realm. The aesthetic refers to a philosophical understanding, reflecting the nature and style of the art employed within the virtual world. Contrary to some perceptions, the Metaverse does not necessarily require the use of mixed reality technologies [16], and the virtual experience can be realized through various mediums, such as VR, 3D, 2D, or even text-based interfaces [30]. Examples of text-based virtual worlds include the renowned board game, Dungeons & Dragons [31], and platforms like

Whatsapp, Facebook, and Instagram, which Meta regards as integral parts of their Metaverse. The choice between VR, 3D, 2D, text-based, or a combination of these mediums should be driven by considerations of cognitive expenditure and the duration of activity. From our observations, VR or 3D is suggested for small group discussions or tutorials where students can focus more completely on full-bodied interactions. Conversely, for larger classes, a 2D environment with strategic camera settings is recommended to avoid potential cognitive overload, as recognizing numerous 3D avatars simultaneously can be challenging. Text-based interfaces can function as integral tools within the virtual world, serving as forums or message boards. Discord, as a popular platform for information exchange, exemplifies this utility. Furthermore, the physical layer necessitates a precise orchestration of object behavior, establishing connections between virtual objects, their attributes, and functionalities to guarantee the fulfillment of their predefined roles. If traditional virtual discussion tools, such as camera or screen sharing, fall short of specific requirements, a standard metaverse platform may prove inadequate. Under such circumstances, the development of a customized application becomes essential to craft a more engaging and facilitative discussion environment. Once tailored to the specific needs, the virtual scene can then be integrated into the broader metaverse platform, thereby enriching the overall learning experience.

The first and second layers together delineate the virtual world's configuration, forming what we refer to as a "universe." If such a universe is crafted with educational intent, it becomes an "eduverse." As the applications and needs diversify, we foresee the creation of multiple universes, each tailored to specific purposes. For instance, an eduverse designed to simulate a hospital environment for healthcare education would incorporate unique functions relevant to that setting but not utilized in other eduverses. We term this diverse landscape a "multiverse." To bind these disparate eduverses into a cohesive Eduverse or Edu-Metaverse, the inclusion of a third layer is essential—the economic layer. This layer shapes the system of rewards and exchanges, allowing students to share value across different eduverses. Depending on factors such as security and usability, this economic layer can adopt centralized, decentralized, or hybrid approaches.

The economic layer consists of two essential elements warrant attention. The first one is digital identity. This constitutes the information used by the eduverse to recognize individual students. In our design, we utilize a single sign-on (SSO) system within our university, requiring students to log in with their specific student credentials. The flexibility of the system allows integration with various SSO options, such as Google, Facebook, or WeChat, across different eduverse platforms. There's also the option to permit anonymity, fostering more open peer-to-peer learning during discussions. The second one is digital ownership: Encouraging active student participation in virtual activities is vital. By employing gamification and Play2Earn principles, we generate virtual coins as rewards for students. We have developed specific apps like PolyU WellFit and PolyU GreenCoin to support holistic education. Recent advancements in blockchain technologies have been beneficial in this regard [32]. Our transaction across all apps is linked to PolyChain, a generic blockchain platform, ensuring seamless integration.

In summary, the Eduverse concept within our university encapsulates various virtual worlds and apps, interconnected through SSO and blockchain technology. This multi-faceted structure fosters a vibrant and dynamic learning ecosystem, leveraging the best of both virtual and real-world educational strategies.

3.2 Connect Students by Using Gamification and Constructive Teaching

The metaverse offers a novel learning environment, presenting opportunities to innovate and evolve pedagogical strategies. We contend that the application of the metaverse in education should extend beyond mere replication of traditional lecture halls or tutorial classes. Instead, new approaches that harness the unique capabilities of this virtual space should be explored. In this section, we will delineate the methodology for utilizing the structure of role-playing games (RPG) and the principles of constructivism to design an educational course setting within the metaverse.

Changes in the learning behaviors of our students warrant attention, as these shifts are not isolated from broader cultural phenomena. Some scholars have attributed the resurgence of the metaverse concept among students to the success of films like "Ready Player One" and "Free Guy" [33]. This connection between popular culture and learning suggests that immersive experiences in the metaverse can indeed act as motivational catalysts. We concur with this perspective, recognizing the potential for the metaverse to enrich the educational experience by engaging students in ways that resonate with contemporary cultural trends.

In the context of this project, we seek to introduce a novel pedagogical framework aimed at fostering student engagement and effective learning, namely, Gamified Constructivist Teaching in the Metaverse (GCTM). As the terminology indicates, GCTM is predicated on the integration of two well-established learning theories: gamification and constructivism (see Fig. 2). The main idea of GCTM is to create a new learning environment to support new learning activities and course design.

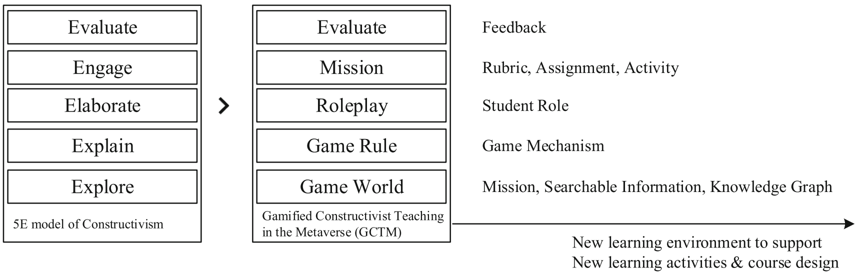


Fig. 3. Gamified Constructivist Teaching in the Metaverse (GCTM) framework

In the Game World Design, which aligns with the explore phase of the 5E model, the traditional systematic lecture approach is replaced with a more engaging strategy. Instead of delivering long talks on specific subjects, educators transform their lecture materials into three distinct components: 1) A well-defined Mission or objective related to the course. 2) Searchable information that students can access and explore; 3) A knowledge graph illustrating the interconnections between various materials; and The details of this setting will be covered in the next session. This innovative approach provides an interactive virtual environment and enables students to take on the role of players, embarking on new and exciting adventures within the Metaverse. By shifting from passive listening to active exploration, students are encouraged to engage with the material, fostering a deeper understanding and more interactive learning experience (Fig. 3).

Game Rule Design corresponds with the explain phase in the 5E model. Once the game world has been established, this stage requires lecturers to furnish an overview of the knowledge and theories to be learned, as well as delineate the connections among these learning contents. To augment the efficiency of this explanation, we advocate the incorporation of game mechanism design. Owing to the widespread success of mobile games, students are often already acquainted with various game mechanisms. Game designers leverage different game mechanisms to encourage players to explore various parts of the game, employing strategies like daily missions, urgent missions, level-up systems, first draws, and more. By performing similar tasks repeatedly, players learn how to navigate the game, developing new habits.

Roleplay Design aligns with the elaborate phase of the 5E model. Once various game mechanisms have been established to systematically convey knowledge, the focus moves beyond crafting open-ended or reflective questions. In this stage, we create specific roles for students within the Metaverse. For instance, students might take on the role of a STEM teacher and design a STEM workshop for primary school students or a game designer in a game company. It aims to transform students who may be passive or struggle with low motivation into engaged and active learners. By immersing them in a simulated environment where they take on professional or creative roles, we encourage them to internalize and apply their knowledge, making the learning experience more tangible and stimulating.

Mission Design corresponds to the engage phase in the 5E model. Educators at this level must create the content for tutorial exercises, assignments, and the final project. Since GCTM utilizes a game system approach, a well-defined rubric must be crafted and presented to the students. This assessment tool should articulate achievement criteria across all components, providing clarity on what is expected of the students. By laying out these expectations, the system helps to identify any knowledge gaps and ensures that students are aware of the learning objectives. The mission design thereby serves as a roadmap, guiding students through their educational journey and fostering a more interactive and engaging learning experience within the Metaverse environment.

Evaluate Design aligns with the evaluate phase in the 5E model. Within this stage, formal assessments are administered to the students using predefined rubrics, while informal assessments may occur during observations of students' discussions and interactions. This design goes beyond merely assessing academic results. It also takes into consideration the learner's experience within the GCTM framework. By incorporating both formal and informal evaluations, educators can gain a comprehensive understanding of not only the students' grasp of the material but also their engagement and satisfaction within the educational environment. This holistic approach helps in creating a more effective and responsive learning experience tailored to the unique demands and opportunities of the virtual world.

3.3 Connection with the Domain Knowledge

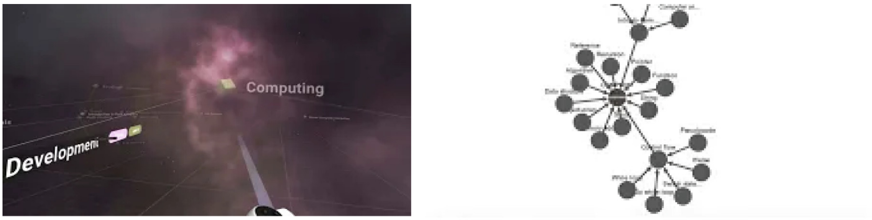
The essence of the game world design component, as well as the GCTM framework, lies in deconstructing domain knowledge into specific missions, searchable information, and a cohesive knowledge graph. This approach enables a more interactive and engaging learning experience, transforming traditional educational content into an adventure within the Metaverse.

Mission: The starting point for lecturers is to define a clear and engaging mission for the students. Examples might include acting as a STEM teacher to develop machine learning courses for primary students or serving as a legislator to evaluate public policy. This mission serves as the goal, with all assignments and activities aligned to support its completion.

Searchable Materials: Instead of traditional slides and lecture recordings, materials must be transformed into bite-sized, searchable resources. This approach enables students to actively explore and discover knowledge within the game world. If videos are the primary learning resources, our research indicates that long videos are not as effective for learning. Shorter video clips, averaging around 8 min and 43 s, have proven more effective in our studies. Technologies like Azure Cognitive Service or Panopto, utilizing AI models for face recognition, translation, computer vision, and speech, can automatically tag and caption videos, enhancing the search experience across a library. Along with this, lecture slides and reading materials should be converted into searchable PDFs. In alignment with the Metaverse concept, materials should consider digital ownership, possibly using blockchain to publicly register content creation.

Systematic Graph: When dealing with vast amounts of searchable knowledge, it's vital to provide an overview that illustrates how different pieces of information are organized and interconnected. This can be likened to skill trees in RPG games, helping students understand their learning journey. While a hierarchical or tree structure is common, some connections between knowledge nodes may create cycles. In our implementation, we use a knowledge graph, a graph-structured presentation of connections between knowledge elements (See Fig. 4). The left-hand side presents a 3D and VR version of our work, showcasing knowledge nodes with slide previews. The right-hand side displays a 2D and web version, providing an easy-to-use interface for students to grasp an overview of the knowledge structure.

In summary, this design fosters an engaging and interactive learning environment by transforming traditional educational content into a Metaverse adventure. Through the careful crafting of missions, searchable materials, and a systematic knowledge graph, educators can encourage active exploration and learning among their students.



(a) VR Knowledge Graph (Link) (b) 2D Knowledge Graph (Link)

Fig. 4. Knowledge graph in GCTM framework

4 Evaluation

The GCTM (Gamified Constructivist Teaching in the Metaverse) framework was implemented in one subject across two engineering classes, with 62 students in one class and over 200 in the other, utilizing the Gather.Town platform. The evaluation of this project was conducted using a combination of tools: a questionnaire designed around the 5I model, the university's standard student feedback questionnaire (SFQ), and a review of the students' academic results at the conclusion of the semester.

The aim of this evaluation was to uncover the relationship between the students' backgrounds and their satisfaction with the GCTM approach. The questions were constructed using a 5-point Likert scale, ranging from 1 for "Strongly Disagree" to 5 for "Strongly Agree." This format facilitated an in-depth understanding of the students' experiences and allowed for a detailed assessment of the GCTM's efficacy in the context of their learning as shown in Table 3.

Our data observation reveals a positive reception of the GCTM setting, with almost all items receiving a rating over four in the larger class. When comparing the satisfaction levels between the virtual world setting and traditional platforms like ZOOM or MS Teams, we found that the virtual world fosters a stronger sense of belonging among students. The interaction rating, which measures both student-to-student and student-to-lecturer interactions, is relatively high. This suggests that the GCTM design enhances opportunities for students to engage with both their lecturers and peers compared to traditional methods. Our observations, as well as student feedback, indicate that when given ample preview and explanation about the tasks and roleplay, students are more willing to exchange ideas and actively participate in discussions. Rather than simply typing answers in the chat, they are more inclined to turn on their cameras and verbally engage, adding a more dynamic and interactive dimension to the learning experience. We have also garnered positive feedback on GCTM, as outlined below. The comments highlight that student enjoyed the enhanced interaction with the lecturer and appreciated this innovative approach:

Comments on Metaverse Platform

- Interesting and effective teaching mode especially using gather town, which is an interesting thing to me
- I like the group assignment that work together in the metaverse learning environment.

Table 3. Student Satisfaction

Item	COMP N = 209	EIE N = 62
Satisfaction (Content)		
Slides	4.00	3.69
Video	4.12	4.03
Satisfaction (Teaching Activities)		
Preview Session	3.86	3.76
Q&A Session	3.84	3.73
Satisfaction (Learning Activities)		
Overall	4.10	4.18
Engagement	3.87	4.00
Reflection	3.82	3.84
Understand	3.91	3.77
Satisfaction (Virtual environment)		
Easy to use	4.32	4.15
Stable	4.14	4.05
More comfortable environment	4.11	3.95
Better sense of belonging	4.11	3.85
Better team building	4.08	3.98
Improve Student-student interaction	4.11	3.95
Improve Student-lecturer interaction	4.06	3.95

COMP = Year 3 students in Department of Computing.
EIE = Year 4 students in Department of Electronic and Information Engineering.

- The topics are interesting and allow me to explore the other side of computer. The idea of collaboration through metaverse is really good and helpful.

Comments on GCTM

- Use independent watching videos to learn, and the teaching method of answering questions in class is very efficient and high-quality
- Watch video in YouTube and consult in class is good teaching method for me, helping me to have a good time management.
- Lecturer already gives me a lot of help on this course, give a very good explanation on the topic by recording video and lectures notes. I am appreciating it.
- Flipped classroom, so that students can use the time on lecture to communicate with teacher.

- He spends so many times to teach us and answer our question one by one.
- The Q&A time during class. I always learn more through asking questions.
- The flipped learning teaching mode.

5 Conclusion

In conclusion, this paper has presented the Gamified Constructivist Teaching in the Metaverse (GCTM) framework, a novel approach to educational engagement in the computer science field. Through the integration of game world design, game rule design, roleplay design, mission design, and evaluate design, GCTM has been shown to enhance student-student interaction, student-lecturer interaction and increase satisfaction levels in the learning environment.

The implementation of GCTM in two engineering classes, utilizing the metaverse platform, revealed promising results. The systematic arrangement of missions, searchable materials, and knowledge graphs created an engaging and immersive educational setting, fostering a stronger sense of belonging and interaction among students.

Furthermore, the positive feedback received from students affirms the success of this approach in enhancing the educational experience. The high ratings across various components of the program, along with students' verbal appreciations, showcase the potential of GCTM in transforming conventional education methodologies.

However, it is essential to recognize that this study represents an initial exploration of the concept. Future research could focus on optimizing various elements of GCTM, implementing it across diverse educational fields, and conducting more extensive evaluations.

Overall, GCTM represents a significant step forward in the integration of gamification and constructivist principles within the virtual world of education. Its success in the observed classes underscores its potential as a versatile and effective tool for modern educators, potentially paving the way for a more engaging, interactive, and learner-centered educational landscape.

Acknowledgements. This work was supported in full by the Hong Kong Polytechnic University, Project of Strategic Importance (project number: P0036846) and Teaching Development Grant (project number: TDG22–25/VTL-11).

References

1. Rashid, S., Yadav, S.S.: Impact of Covid-19 pandemic on higher education and research. Indian J. Hum. Develop. **14**(2), 340–343 (2020Aug)
2. Paudel, P.: Online education: Benefits, challenges and strategies during and after COVID-19 in higher education. Int. J. Stud. Educ. (IJonSE). **3**(2), (2021)
3. Ferrara, J.D.: Teaching a large-scale crystallography school with Zoom Webinar. Struct. Dyn. **8**(1) (2021)
4. Martin, L., Tapp, D.: Teaching with teams: an introduction to teaching an undergraduate law module using microsoft teams. Innov. Pract. High. Educ. J. **3**(3), 58–66 (2019Apr 7)
5. Oakeshott, M.: The idea of a university. In: Academic ethics 2017 May 15 (pp. 9–16). Routledge

6. Cope, M.: The seven Cs of consulting. Pearson UK; 2010 Oct 5 (2010)
7. May, G.L.: Collaborative learning techniques:a handbook for college faculty. Pers. Psychol. **58**(4), 1097 (2005Dec 1)
8. Chen, P.Q., Ng, P.H.: College students experience of service learning in teaching STEM workshop Under COVID-19. In: 2022 IEEE International Conference on Teaching, Assessment and Learning for Engineering (TALE) 2022 Dec 4 (pp. 404–409). IEEE (2022)
9. Sin, Z.P.: Towards a metaverse of knowledge: a constructivist proposition on immersive learning and visual literacy. In: International Conference on Web-Based Learning 2022 Nov 21 (pp. 214–225). Cham: Springer International Publishing
10. Sin, Z.P.: Towards an edu-metaverse of knowledge: immersive exploration of university courses. IEEE Trans. Learn. Technol. (2023)
11. Peter, H.F., et al.: Implement virtual reality tour in blended learning. In: Richard Chen Li, Simon K. S. Cheung, Peter H. F. Ng, Leung-Pun Wong, Fu Lee Wang, (ed.) Blended Learning: Engaging Students in the New Normal Era: 15th International Conference, ICBL 2022, Hong Kong, China, July 19–22, 2022, Proceedings, pp. 74–84. Springer International Publishing, Cham (2022). https://doi.org/10.1007/978-3-031-08939-8_7
12. Qing, L. et al.: KCUBE: A Knowledge Graph University Curriculum Framework for Student Advising and Career Planning. In: Richard Chen Li, Simon K. S. Cheung, Peter H. F. Ng, Leung-Pun Wong, Fu Lee Wang, (ed.) Blended Learning: Engaging Students in the New Normal Era: 15th International Conference, ICBL 2022, Hong Kong, China, July 19–22, 2022, Proceedings, pp. 358–369. Springer International Publishing, Cham (2022). https://doi.org/10.1007/978-3-031-08939-8_31
13. Radoff, J.: The metaverse value-chain. Medium (2021)
14. Brian, J.: Why Blockchain Isn't a Part of Meta's Metaverse – Yet. Spiceworks (2022)
15. Mudrick Capital Management. Remark Holdings Receives $30 Million of Debt Financing from Mudrick Capital (2021)
16. Park, S.M., Kim, Y.G.: A metaverse: taxonomy, components, applications, and open challenges. IEEE access. **4**(10), 4209–4251 (2022Jan)
17. Ng, P.H.: Introducing the practices for adopting the constructivist teaching in game engineering. In: 2017 IEEE Global Engineering Education Conference (EDUCON) 2017 Apr 25, pp. 1636–1643. IEEE (2017)
18. Hamari, J., Koivisto, J., Sarsa, H.: Does gamification work?--a literature review of empirical studies on gamification. In: 2014 47th Hawaii international conference on system sciences 2014 Jan 6, pp. 3025–3034. IEEE (2014)
19. Chou, Y.K.: Actionable gamification: Beyond points, badges, and leaderboards. Packt Publishing Ltd. (2019)
20. Beilin, H.: Piaget's enduring contribution to developmental psychology. Dev. Psychol. **28**(2), 191 (1992Mar)
21. Harlow, S., Cummings, R., Aberasturi, S.M.: Karl Popper and Jean Piaget: a rationale for constructivism. In: The Educational Forum 2007 Mar 31 (Vol. 71, No. 1, pp. 41–48). Taylor & Francis Group (2007)
22. Ng, P.H., Chiu, M.W.: Where programming skills meet the social needs. Service-Learning for Youth Leadership: The Case of Hong Kong, pp. 165–84 (2019)
23. Bybee, R.W.: What is STEM education?. Science. 2010 Aug 27 **329**(5995), 996 (2010)
24. Fazelian, P., Soraghi, S.: The effect of 5E instructional design model on learning and retention of sciences for middle class students. Procedia Soc. Behav. Sci. **1**(5), 140–143 (2010Jan)
25. Jobrack, B.: The 5E Instructional Model: Engage Explore Explain Evaluate Extend. From Science, Technology, Engineering and Mathematics. https://www.mnheonline.com/secondaryscience.pdf. (2013)
26. Hao, J., Yan, Y., Gong, L., Wang, G., Lin, J.: Knowledge map-based method for domain knowledge browsing. Decis. Support. Syst. **1**(61), 106–114 (2014)

27. Janowicz, K., et al.: Know, Know Where, KnowWhereGraph: a densely connected, cross-domain knowledge graph and geo-enrichment service stack for applications in environmental intelligence. AI Mag. **43**(1), 30–39 (2022)
28. Lin, J., Zhao, Y., Huang, W., Liu, C., Pu, H.: Domain knowledge graph-based research progress of knowledge representation. Neural Comput. Appl. **33**, 681–690 (2021)
29. Fryer, L.K., Shum, A., Lee, A., Lau, P.: Mapping students' interest in a new domain: connecting prior knowledge, interest, and self-efficacy with interesting tasks and a lasting desire to reengage. Learn. Instr. **1**(75), 101493 (2021)
30. Bartle RA. Designing virtual worlds. New Riders; 2004
31. Dionisio, J.D., Iii, W.G., Gilbert, R.: 3D virtual worlds and the metaverse: current status and future possibilities. ACM Comput. Surv. (CSUR). **45**(3), 1–38 (2013)
32. Berg, C., Davidson, S., Potts, J.: Blockchain technology as economic infrastructure: revisiting the electronic markets hypothesis. Front. Blockchain **4**(2), 22 (2019)
33. Duan, H., Li, J., Fan, S., Lin, Z., Wu, X., Cai, W.: Metaverse for social good: a university campus prototype. In: Proceedings of the 29th ACM International Conference on Multimedia 2021, pp. 153–16 (2021)

Exploring the Transformative Potential of Virtual Reality in History Education: A Scoping Review

Yalan Zhang[1], Ali Ahmed[2,3](✉), Hai-Ning Liang[4], and Nilufar Baghaei[1]

[1] School of Electrical Engineering and Computer Science, The University of Queensland, Brisbane, Australia
n.baghaei@uq.edu.au

[2] Victoria University of Wellington, Wellington, New Zealand
ali.ahmed@vuw.ac.nz

[3] Department of Computer Science, Faculty of Computers and Artificial Intelligence, Cairo University, Giza, Egypt
a.ahmed@fci-cu.edu.eg

[4] Xi'an Jiaotong-Liverpool University, Suzhou, China
HaiNing.Liang@xjtlu.edu.cn

Abstract. This paper reviews serious games and Virtual Reality (VR) integration into teaching and learning. Serious games are educational video games designed to engage users and improve learning outcomes. VR technology provides an immersive virtual space that enhances learning experiences through visual, auditory, and tactile modes. Mobile games offer portability, allowing students to access educational content and learn outside traditional in-person classes. The integration of serious games and VR has the potential to provide immersive and engaging learning experiences. The paper specifically focuses on the effects of VR and mobile games in history education. It explores how these technologies enhance history learning, provide personalised and smart learning experiences, and improve learning enthusiasm and academic performance compared to traditional teaching methods. The paper follows a scoping review methodology to explore the potential of VR and mobile educational games to facilitate learning history and provide smart or personalised history learning. The findings indicate improved learning enthusiasm and academic performance compared to traditional teaching methods.

Keywords: Virtual Reality · History Education · Scoping Review · Serious Games · Games in Education

1 Introduction

This scoping review aims to systematically explore and synthesise existing literature on the impact and challenges of incorporating Virtual Reality technology into history education, to identify gaps and future directions for research in

H. Xie et al. (Eds.): ICWL 2023, LNCS 14409, pp. 107–121, 2023.
https://doi.org/10.1007/978-981-99-8385-8_9

this emerging interdisciplinary field. Studying history is an important human social practice activity that captures the past and allows individuals to reflect on their identity and cultural heritage. However, it is a fact that history cannot be repeated or experimented upon, making it challenging to experience firsthand [23]. Therefore, history teaching relies heavily on teaching aids such as words, images, and movies to visualise abstract concepts and materialise general information [12]. Language descriptions, in particular, play a significant role in transforming historical knowledge into vivid pictures of life, enabling students to indirectly know or perceive history [5,15]. This approach creates an ideal creative space for developing students' imagination and cultivating their image thinking, which allows them to understand the past more profoundly and engage in critical analysis and interpretation [15]. History is a subject that encompasses a broad range of knowledge, information, and events, which can make it challenging for students to engage and maintain interest [23]. The complex background and numerous characters involved in historical events can make it difficult for students to comprehend and contextualise the presented information. Unfortunately, this can lead to a lack of motivation to learn history. Therefore, it is essential to systematically and innovatively approach history teaching to enhance student learning experiences and improve their engagement and motivation [36]. Systematic teaching is equally essential in history teaching. Teachers should organise their teaching materials and methods to ensure that students can follow the logical sequence of historical events and understand the relationship between them. This approach can help students better understand the historical narrative and make it easier to remember and retain information [23]. Innovative teaching methods can help to make history lessons more exciting and engaging for students. Teachers can use multimedia, simulations, and interactive activities to enhance students' understanding of historical events and promote their critical thinking skills [36]. One of those innovative ways of teaching history is the use of virtual reality (VR).

Furthermore, using VR technology in history teaching can enhance students' understanding of historical events by creating immersive and interactive learning experiences. VR technology can simulate historical environments, allowing students to experience history firsthand and understand the context in which events occurred. For instance, students can visit historical sites and monuments virtually, allowing them to observe and interact with objects and artefacts that may not be readily available to them in their physical learning environment. This approach can help students better understand the cultural significance of these objects and how they relate to historical events. In addition to enhancing students' understanding of history, VR technology can also help to develop their critical thinking and analytical skills. Students can explore historical events from different perspectives, enabling them to develop empathy and understand the complexities and nuances of historical events [22]. They can also engage in problem-solving activities and decision-making scenarios, allowing them to apply their knowledge of historical events to real-world situations. Moreover, VR technology can create an ideal environment for project-based learning, where students can collaborate to develop their historical narratives and explore different historical events and scenarios.

Despite the potential benefits of VR technology in history teaching, some challenges must be addressed. The cost of implementing VR technology in the classroom is a significant challenge for many schools and educational institutions. While the cost of VR technology has decreased in recent years, it remains relatively expensive compared to traditional teaching methods. Additionally, VR technology requires specific hardware and software, which may not be readily available in all schools or require significant investments. This can limit the accessibility of VR technology in history teaching, particularly for schools with limited resources. Another challenge is the lack of standardisation and guidelines for developing VR content for educational purposes. Therefore, developing a comprehensive framework for VR technology's practical use in history teaching is essential to ensure that it is accessible and affordable to all students and is aligned with curriculum standards and educational goals.

2 Literature Survey

VR technology has the potential to revolutionise teaching and learning by creating immersive and engaging learning experiences for students. One of the critical benefits of VR technology is that it allows teachers to use 3D models to demonstrate the content to the students, enabling them to have a more intuitive understanding of the teaching content and realise abstract concepts [10]. Moreover, applying VR technology in teaching activities can go beyond creating virtual teachers, virtual labs, and virtual classrooms. It can also create virtual environments where students can explore and learn about related issues [1,30]. This way, complex historical facts can be visualised in a virtual environment, enabling students to obtain more information through self-experience in virtual vision and virtual feeling. This facilitates student behaviour and enthusiasm in the learning process and enhances their retention of historical knowledge [21,29]. Using VR technology, students can step into historical scenes and interact with them as if they were there. This provides them with a unique opportunity to learn about historical events in a way that textbooks and other teaching materials cannot provide. For example, students can virtually visit historical sites, walk around, and explore architecture and other historical details. They can also see how people dressed, interacted with each other, and lived in that era, which can help them better understand historical events. The popularity of mobile games, particularly among teenagers, has increased due to the rapid growth of mobile media. Mobile devices provide a convenient and portable platform for mobile learning, enabling learners to access educational resources anytime, anywhere. The flexibility offered by mobile technology allows for personalised learning experiences and greater control over the learning process. Research suggests that mobile games can enhance learning motivation and engagement, particularly in history education [11]. Mobile games' immersive and engaging nature contributes to increased student interest and improved learning outcomes. Additionally, mobile games provide a context-specific learning environment, allowing learners to play and learn in similar settings. Overall, mobile games can potentially improve the effectiveness and efficiency of education.

The work in [32] discusses a meta-analysis of 21 experimental studies on the impact of virtual reality (VR) on learning outcomes in K-6 students. The studies show that VR enhances learning outcomes compared to control conditions, with an effect size of 0.64. The effect is even more significant when immersive VR is used, with an effect size of 1.11, compared to semi-immersive and non-immersive systems. The positive effect of VR does not depend on educational level or knowledge domain. Short interventions, less than two hours, were more effective than longer ones. While meta-analysis provides valuable insights into the impact of virtual reality on learning outcomes, there are a few potential criticisms of the work, such as the limited sample size. The meta-analysis only included 21 experimental studies, which may limit the generalisation of the findings.

The work in [7] explores the use of Immersive Virtual Reality (IVR) technology as a remote teaching tool for educational institutions, focusing on teaching architectural history using the Pantheon in Rome as a test case. The study involved two assessments with 57 and 68 students, respectively. It evaluated five independent variables related to learning about architecture, history, sense of presence in VR, structural realism, and comparison to in-class learning. The findings suggest that IVR technology in teaching architectural history provides an excellent learning experience, allowing students to accurately gauge, recognise, and appreciate virtual spaces' 3D aspects, size, and proportion. While the sample size is better than that in [32], Generalisability is a concern. The study only focuses on one test case, the Pantheon in Rome, which may limit the Generalisability of the findings to other historical sites or architectural styles. Further research would be necessary to determine the effectiveness of IVR for teaching a more comprehensive range of architectural history topics.

The work in [14] found that immersive technology, specifically Virtual Reality (VR), positively correlates with positive outcomes such as increased excitement for learning, motivation, deeper learning, and long-term retention. Students reported deeper learning with VR compared to other media or engagement strategies. The study recommends providing professional development for faculty and students to lower the barriers to entry, being flexible in designing learning options for students, making learning exciting and building community, and adopting VR technology for more assignments or in their classrooms. The study also found that students and faculty members are generally open to using VR technology, and its adoption can benefit education.

The work in [24] explores the potential of gamification in historical education and its impact on developing critical professional competencies. An experimental study was conducted with forty students who participated in a 3-month online learning course called "Technology for Constructing Historical Interpretations" using gamified learning techniques. The study found that gamification led to the acquisition of practical knowledge and allowed learners to acquire skills highly relevant to professionals of the 21st century. The results suggest that gamification in history education positively affects the evolutionary development of a personality adaptive to the socio-economic conditions of the 21st century. One limitation of this paper is that no quantitative assessments confirm the effectiveness of gamified learning in history education. As a result, future research should focus on developing assessment tools to demonstrate the effectiveness of gamified educational practices.

Finally, the work in [6] used the Technology Acceptance Model (TAM) to analyse students' acceptance and intention to use AR and VR technologies in university teaching. They found that using AR and VR objects in teaching aroused great interest among the students, and they expressed a high degree of acceptance of the technologies utilised. AR and VR objects' design and usability were also noted to be essential for their acceptance. The study found no significant differences in acceptance between the AR and VR objects used, suggesting that students perceived participating in a mixed reality experience, drawing no distinctions between both resources. However, the authors note that the study's small sample size and absence of a control group are limitations. The concern about this study is how inclusive the findings are. The study sample is not large enough to generalise the findings, and it needs to include more diverse participants.

3 Research Methodology

3.1 Research Objectives

In this study, the authors followed a scoping review methodology proposed in [2] to explore the potential of VR and mobile educational games to facilitate learning history and provide smart or personalised history learning. A scoping review aims to provide an overview of the existing research literature on a particular topic, research question or concept without necessarily attempting to answer a specific research question [2]. Scoping reviews are often used to map out the breadth and depth of a particular area of research, identify gaps in the existing literature, and help to inform future research agendas [18,31]. Unlike a systematic review or a meta-analysis, a scoping review does not usually assess the quality of the studies included, nor does it attempt to synthesise the results of those studies. Instead, it focuses on identifying the key concepts, definitions, and methods used in the literature and presenting a descriptive summary of the available evidence. The scoping review process typically involves a systematic search of the literature, screening of articles for relevance, and data extraction and synthesis. The results of a scoping review may be presented in various formats, including tables, diagrams, and narrative summaries.

The study aims to investigate (1) what ways can the use of VR and mobile educational games facilitate learning history? and (2) Do VR and mobile educational games provide smart or personalised history learning? VR technology has been valued in teaching for its ability to create a more immersive and engaging learning environment. In history education, VR technology and mobile games have significant educational value due to the vast number of knowledge points on the subject and the difficulty of vividly imparting knowledge of buildings and events to students. To answer the first research question, the study will investigate the methodology of applying VR technology and mobile games in history education, the types of interventions used, and the interactions between the participants during the intervention. The second research question will explore the application and limitations of VR technology and mobile games in the smart learning of history education. The study will examine how VR technology and mobile games can provide personalised history learning experiences and enhance the efficiency and effectiveness of history learning.

In this study, the authors employed a systematic literature review approach to screen and select relevant literature using publication year and keywords as criteria. This enabled them to examine the field of study comprehensively, identify recent publications, and focus on the most relevant sources. The literature analysis allowed them to understand the field's development and identify research status, hotspots, and future directions. This knowledge can aid researchers in developing proposals and finding potential areas for collaboration. The approach provided a clear overview of the field's development and facilitated informed decisions regarding future research.

This study's search for relevant literature was mainly conducted using Google Scholar. The search terms for the two groups of search categories were determined by combining subject terms with free words. For VR, the search terms included "History AND Game", "History AND Education", and "History AND Learning". For mobile games, the search term was "History". The increased availability and affordability of VR technology since 2014 has led to its widespread adoption in various fields, including education, resulting in a surge of interest and research on its application in educational settings. Thus, the time range for this study was limited to 2018 to 2022. The search strings were: "VR: History AND Game, History AND Education, History AND Learning" and "Mobile game: History" using Google Scholar. Google Scholar is a comprehensive tool to search multiple disciplines (i.e. education, gamification, etc.) that requires a broad search across various fields.

To ensure the inclusion of only relevant literature, the study applied exclusion criteria, including removing duplicates and articles written in languages other than English, excluding non-national history topics like architectural or art history, and using advanced search options on Google Scholar to filter articles by title. The search strategy found 34 articles, from which 29 were screened for relevance, and 12 were ultimately excluded for not being pertinent to the topic. The next step involved assessing the eligibility of the remaining 17 articles using the defined criteria. A full-text review was conducted, and five articles were removed, leaving 12 articles for the scoping review. The selection process for this study is presented in Fig. 1, which provides a visual representation of the search strategy and selection criteria used to identify relevant articles for the scoping review. This process ensured that only articles that met the study's eligibility criteria were included in the final analysis.

Table 1 provides a detailed breakdown of the study selection process for different keywords. The table illustrates the number of articles found for each keyword, the number of duplicates, and the number of articles excluded based on their irrelevance or failure to meet the eligibility criteria.

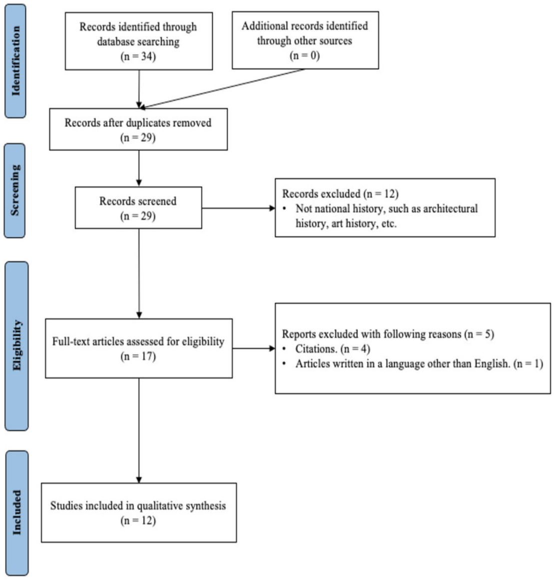


Fig. 1. Literature Screening and Selection flowchart

The data extraction process is a crucial step in systematic reviews of educational games, involving downloading the full text of selected articles and extracting relevant information such as publication year, history content, participant demographics, data collection methods, experiment methodology, and critical findings. This data is essential for understanding the effectiveness of educational games in promoting learning and identifying patterns and trends in game design, evaluation, and effectiveness across different contexts. The extracted data from systematic reviews of educational games can serve multiple purposes: guiding the development of experimental protocols, informing the design and evaluation of future games, and ensuring a systematic and rigorous approach to promote learning and enhance the quality of educational games. Researchers can develop robust protocols for evaluating these variables by identifying key variables associated with successful outcomes. Critical findings about users' games and learning experiences also help design more engaging and effective games. This comprehensive approach contributes to advancing educational game design and evaluation, ultimately improving learning outcomes.

4 Results Discussion

After the screening, 12 articles were chosen for a scoping review. The selected articles were published over the past 5 years. The selected articles' analysis

Table 1. Article Inclusion and Exclusion

Search Terms	Total Articles (%)	
	Collected	Excluded
VR AND History AND Game	6 (23.08)	2 (11.76)
VR AND History AND Education	6 (23.08)	3 (17.65)
VR AND History AND Learning	14 (53.84)	7 (41.12)
Sub-Total	26 (100)	17 (100)
Mobile game AND History	8 (100)	4 (100)
Sub-Total	8 (100)	5 (50)
Total	34 (100)	22 (61.76)

showed no clear trend in the number of relevant articles published over the past five years. In most years, a similar number of relevant articles were published. In 2018, 4 relevant articles (33%) were published, while in 2019, 3 relevant articles (25%) were published. In 2020, only one relevant article (8%) was published. However, in 2021, the number of relevant articles published increased to three again (25%). Overall, based on these data, it can be concluded that there is no significant upward or downward trend in the number of relevant articles published over the past five years.

The total number of participants across all studies included in the analysis is 598. Figure 2 depicts the participants' educational backgrounds. Out of these, 495 participants have age information available. Among those with age information, 153 participants (31%) fall within the age range of 8–12 years old, while 342 participants (69%) fall within the age range of 13–34 years old. The studies also included participants from diverse backgrounds, from primary school to undergraduate level. Out of the 500 participants with information on educational level, 171 (44%) were elementary school students, 50 (13%) were junior high school students, 66 (17%) were high school students, and 105 (26%) were college students. Overall, the findings suggest that the studies included in the analysis recruited participants from a wide range of ages and educational backgrounds, providing a diverse sample for the experiment and assessment.

Participant demographics are essential to any study as they provide a better understanding of the study sample and its characteristics. In the present study, as Table 2 shows, gender information was not reported in nearly half of the studies included in the analysis, which limits the ability to draw accurate conclusions about gender differences in the study findings. However, among the studies that did provide gender information, 143 participants were male (42%), and 197 were female (58%). This unequal gender distribution may impact the study findings and should be considered when interpreting the results. In terms of the distribution of the total number of participants across the included studies, it was found that 7 out of 11 experiments had a sample size of 5–50 participants (64%), three experiments had a sample size of 66–98 participants (27%), and 1

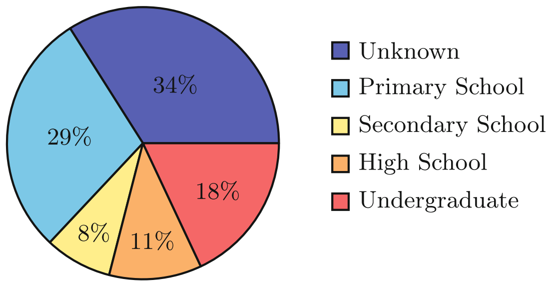


Fig. 2. Pie chart showing the educational background of participants

experiment had a sample size of 160 participants (9%). This variation in sample size across studies may impact the study findings and should be considered when interpreting the results.

Table 2. History content involved and participants' data

ID	History content	Participants (Male, Female)
[16]	Greek myth	28 primary school students
[26]	Madurese history and alphabet	29 3-rd grade students, 5 teachers
[35]	Kaaba related to Islam History	25 undergraduate students (M = 12, F = 13)
[33]	The Roman Civilisation (Augusta Emerita)	98 pupils aged 9–10, coursing fourth year of Primary Education (M = 53, F = 46)
[28]	History of the Netherlands	16 kids from class 7 and 8 of the Mariaschool in Beuningen (M = 9, F = 7)
[25]	Battles between the Japanese and Australian forces over the Australian territory, Papua, during World War II	80 undergraduate students from a university in Southern California (M = 19, F = 61)
[27]	History and life in Ancient Rome	50 first-year Secondary School students (M = 27, F = 23)
[8]	History of the Ningbo Sanjiangkou Site and heritage objects	160 undistinguished participants
[13]	–	–
[20]	History of the town of Kemijärvi, Northern Finland	5 children (M = 0, F = 5)
[17]	King Ludwig II	36 teenagers who stayed close to the starting point of the game (Max-Joseph-Platz; Munich)
[9]	Chinese history (Qing dynasty)	66 high school students (M = 24, F = 42)

4.1 Investigating the Research Questions

How do the use of VR and mobile educational games facilitate history learning? All reviewed articles' experiment results and primary findings have been compiled and summarised in Table 3. According to Table 3, most students who participated in the VR game-based learning experience had a positive overall experience. The students perceived the VR games used in these studies as simple, practical, and innovative. Additionally, younger participants, in particular, were found to be excited about the prospect of learning history through VR games. They showed great interest and curiosity in the game elements and characters, as highlighted in [8] and [20]. This level of engagement stimulates their enthusiasm for learning and helps to keep them motivated throughout the learning process.

Moreover, the results of various studies indicate that students who use VR or mobile games for history learning are more effective in acquiring content covered in the lesson and demonstrate a statistically significant improvement in their test scores compared to students taught through traditional methods based on a coursebook. Studies in [27] and [17] highlight this finding. Participants in these studies noted that using VR technologies for history learning removed negative biases they may have had towards history courses, making the learning process more enjoyable, as also noted in the study in [35]. Furthermore, VR technology benefits students' learning experience and brings convenience to the broader public. This is particularly true for individuals who face physical or financial barriers to accessing historical locations. Elderly people or people with walking disabilities may find it challenging to visit destinations physically, but VR technology allows them to explore and experience such locations virtually. The study in [35] noted this aspect, where participants acknowledged that the technology offered opportunities to visit destinations they could not otherwise go to.

Do VR and mobile educational games offer personalised and smart history learning experiences? To answer whether smart or personalised learning was provided to participants in the studies reviewed, experiment methodologies were extracted and summarised in Table 4. However, upon further analysis, it was found that none of the 12 articles in the scoping review offered personalised learning. To explore this topic further, a similar search strategy was used, focusing on the keywords "personalised learning", "VR", and "mobile games" in recent articles on Google Scholar. The search results revealed that most personalised services are geared towards providing users with recommendations, such as advertisements or rehabilitation training. Only a few articles were found related to personalised history learning in education.

The study in [3] introduced a mobile app called JuJu English Vocabulary to improve learners' vocabulary acquisition using the Keller Plan Personalised system of instruction theory. The app was tested in a quasi-experimental study

Table 3. Summary of studies' findings

ID	Summary of findings
[16]	The VR game was considered simple, practical, predictable, and appropriate, with a stimulating, innovative, and valuable game environment.
[26]	The feasibility study results showed the feasibility level in effectiveness, efficiency, and attractiveness. The game also functioned as a strengthening of the learning material for students.
[35]	Participants felt a part of the environment and felt present in the exact location. Besides, participants also stated that VR technology-facilitated learning by providing enjoyable learning environments and interesting content.
[33]	Results showed that students who used VR for history learning had a statistically significant improvement in their test scores compared to those taught through the traditional method based on a coursebook.
[28]	–
[25]	Immersive environments may create excessive positive emotions and distraction, distracting the learning engagement in appropriate cognitive processing of the information. Learners performed worse after viewing the immersive VR (IVR) lesson than the video lesson on transfer questions.
[27]	AR app group has more excellent grades and more high-achieving students. The performance statistics show that almost half of the students in this group have absorbed most of the concepts they were taught.
[8]	Both age and education level impacted the gaming experience. Younger participants were relatively quick to adapt to VR technology. Participants with relatively high levels of education pay more attention to the potential in learning the game's history but learn less during the process.
[13]	-
[20]	Game elements and details that make game characters interesting are important to keep children engaged. The results of the end quiz showed that participants remembered information about historical figures well.
[17]	Participants who used a combination of the smartwatch and a mobile game answered most questions correctly. Most participants believed the mobile game and the smartwatch were easy to understand and use.
[9]	Results indicated a significant improvement in students' performance in history learning after playing the mobile game. Additionally, after the gameplay, students learned more about the relations between characters and events.

involving 60 participants, and the results showed improved performance in the post-test for the experiment group compared to the control group. However, the article did not explicitly mention the advantages and benefits of personalised services. Another work in [34] proposed an architecture for an adaptive recommendation model of online learning resources. Through data analysis, the author

Table 4. Summary of studies' experiment methodology

ID	Method	Participants' task/interaction
[16]	Students played "Trials of the Acropolis" after an introduction to the game and technology, wearing cardboard and earphones	Game lasted 15 mins, with 3 required trials.
[26]	Users explored Sumenep Palace Museum's historical relics in "Halo Madura" game, using VR Box and controller to move around.	Users can read information about historical objects in Madura upon approaching them.
[35]	Participants had flexible study opportunities using VR glasses and rotating chairs, with no time limits.	Participants learned about Kaaba in Islamic history education using audio and visual content. They interacted with information points and walked around to obtain audio information.
[33]	Pupils in the VR group observed examples of Roman Augusta Emerita's houses using VR glasses and VirTimePlace app, while the control group used the coursebook with images on paper. The session lasted 45 min	Participants virtually toured Roman Augusta Emerita to observe its representative houses and buildings.
[28]	Participants received Oculus Go headset with an installed application and used it while communicating with the supervisor	TParticipants verbally communicated their location in the application to the supervisor by reading their surroundings aloud.
[25]	The IVR group used a hand-held controller to experience an academic history lesson called Kokoda VR, while the video group used a mouse. Both groups had a 22-min session.	The IVR group used a hand-held controller to interact with objects, walk in a small space and view objects from different angles, while the video group used a mouse and had a stationary perspective without interaction.
[27]	AR group: 60-min lecture + 30-min AR game session. Traditional group: 60-min lecture + 30-min independent study.	Participants could design a city or use a pre-designed city during the gameplay, explore existing buildings in first-person mode, and read historical information scrolls in each building.
[8]	The time limit was set to 8 min per group of paired players carrying out their roles (merchant and post officer) after a brief site exploration	Virtual experiments reset for new groups. The merchant starts with default object positions, and the port officer interacts with the environment the merchant has changed.
[13]	-	-
[20]	Students received an interactive prototype of mobile UI via Marvel App. Three 15-min test sessions were conducted	Participants acted as 1920s postmen, delivering invitations to historical figures in Kemijärvi. VR/AR provided character and building info during gameplay.
[17]	Group 1 downloaded the game via a Wi-Fi hotspot and received smartwatches with lights. Group 2 downloaded the game but without smartwatches. Group 3 downloaded a pdf information sheet from a Google Drive link and read it	Participants received info and instructions at each station. Correctly answering questions earned 'Guldens' (15 points). Trophies were collectable, some requiring a specific number of 'Guldens'.
[9]	The experiment lasted 90 min (10 min intro/pretest, 50 min gameplay, 10 min posttest, 10 min questionnaire). 8-in mobile devices had the game system preloaded	The mobile game had 5 events and 50 figures. Participants entered the number of figures in each event into limited answer slots

found that continuous personalised and fragmented online learning improved learners' levels and cognitive abilities to varying degrees. The article concluded that online learning based on mobile devices is feasible and effective when using the construction model of personalised fragmented resource recommendation scenarios.

As for VR, very few articles were found related to personalised learning. The work in [19] introduced software for training users to inspect construction sites for potential hazards using personalised VR training scenarios synthesised through an optimisation method based on user training preferences and target training time. User research validation found that personalised guided VR Training can improve the user's construction hazard inspection skills more effectively. In conclusion, personalised learning is beneficial in education. However, personalised services are primarily used in business, and their application in education, especially in educational games, is still limited. Although some articles are related to personalised learning in education, further research is needed to explore its potential benefits and drawbacks.

To better understand the participants in these studies, an exploratory data analysis was conducted to visualise their demographics, such as their gender, age, and educational background. Statistics showed that in experiments with known participant age information, most average ages were between 13 and 34 years old (69.1%). The age range of all participants was 8–34 years, indicating a lack of research on the older population. The exploratory data analysis also revealed almost no gender bias across the studies, as the gender ratio of males (143) to females (197) was almost the same. Interestingly, elementary (171) and undergraduate (105) students accounted for 46.2% of all participants, highlighting the need for a more diverse sample set for conclusive proof of the presented results. In 2008, the Australian government signed the Melbourne Declaration on Education, which stated that the educational goals included "promote personalised learning that aims to fulfill the diverse capabilities of each young Australian" and "all young Australians become successful learners, confident and creative individuals..." [4]. This document illustrates that personalised education is a significant area of education reform. However, due to the complexity of personalised services, personalised education has not been applied to VR and mobile educational games on a large scale, especially in the history curriculum. Therefore, existing games cannot provide students with personalised content, timely support, and feedback.

5 Conclusion and Future Work

Most studies in this literary analysis utilised design concepts, theories, and models to evaluate VR and mobile games, demonstrating positive effects on history learning. However, one experiment in [25] indicated lower cognitive engagement and academic performance with immersive VR technology. These inconsistent findings highlight the early stage of VR's application in history education, prompting further research and critical thinking. The experiments reviewed had a limited sample size (598 participants across 11 experiments), with a significant number of experiments (54.5%) involving fewer than 36 participants. This indicates the need for stronger reliability and consistency in the conclusions drawn. Further improvements and development are necessary in the application of VR and mobile games in history education, including determining appropriate game modes for different historical content.

When interpreting the results of this review, it's important to consider limitations for future work. The search strategy only included articles from Google Scholar, potentially excluding relevant articles from other databases. Additionally, the search was limited to keywords in article titles, potentially missing relevant articles. Future reviews should expand the search to include additional databases and broaden the search terms.

References

1. Aljuhani, A., Alghamdi, R., Aldakhil, A.: Virtual reality applications in education. Int. J. Emerg. Technol. Learn. (iJET) **13**(12), 122–134 (2018)
2. Arksey, H., O'Malley, L.: Scoping studies: towards a methodological framework. Int. J. Soc. Res. Methodol. **8**(1), 19–32 (2005)
3. Arumugam, R., Md Noor, N.: Mobile apps based on Keller personalized system of instruction to promote English vocabulary acquisition. Int. J. Interact. Mob. Technol. **15**(23), 4–17 (2021)
4. Barr, A., et al.: Melbourne Declaration on Educational Goals for Young Australians. Ministerial Council on Education, Employment, Training and Youth Affairs (2008)
5. Bilbokaitė, R.: Effect of computer based visualization on students' cognitive processes in education process. In: SOCIETY. INTEGRATION. EDUCATION. Proceedings of the International Scientific Conference, vol. 4, pp. 349–357 (2015)
6. Cabero-Almenara, J., Llorente-Cejudo, C., Martinez-Roig, R.: The use of mixed, augmented and virtual reality in history of art teaching: a case study. Educ. Sci. **10**(10), 286 (2020)
7. Chan, C.S., Bogdanovic, J., Kalivarapu, V.: Applying immersive virtual reality for remote teaching architectural history. Educ. Inf. Technol. **27**, 4365–4397 (2022)
8. Cheng, D., Ch'ng, E.: Potentials for learning history through role-playing in virtual reality. In: Association of Critical Heritage Studies 5th Biennial Conference, ACHS 2020 Futures, pp. 1–18 (2020)
9. Chou, Y.S., Hou, H.T., Su, C.L., Chang, K.E.: Designing and evaluating a mobile educational game "void broken 2.0" for history instruction. In: 27th International Conference on Computers in Education, ICCE 2019, pp. 11–18. Asia-Pacific Society for Computers in Education (2019)
10. Elmqaddem, M.: Impact of virtual reality technology on teaching and learning. Int. J. Educ. Pract. **7**(8), 115–123 (2019)
11. Fendt, M.W., Ames, E.: Using learning games to teach Texas civil war history to public middle school students. In: 2019 IEEE Conference on Games (CoG), pp. 1–4. IEEE (2019)
12. Gamar, M.M., Tati, A.D.R., et al.: Utilization of ICT-based learning media in local history learning. J. Phys. Conf. Ser. **1764**(1), 012079 (2021)
13. Homan, J.C.: The possibilities of virtual reality-goggles within the teaching and learning of history. Yesterday Today **20**, 175–185 (2018)
14. Hutson, J., Olsen, T.: Virtual reality and art history: a case study of digital humanities and immersive learning environments. J. High. Educ. Theor. Pract. **22**(2), 50–65 (2022)
15. Ismail, M.E., Othman, H., Amiruddin, M.H., Ariffin, A.: The use of animation video in teaching to enhance the imagination and visualization of student in engineering drawing. IOP Conf. Ser. Mater. Sci. Eng. **203**, 012023 (2017)
16. Kazanidis, I., Palaigeorgiou, G., Chintiadis, P., Tsinakos, A.: A pilot evaluation of a virtual reality educational game for history learning. In: European Conference on e-Learning, pp. 245–253. Academic Conferences International Limited (2018)
17. Leo, M.: Combining a location-based mobile game and a connected wearable for history learning purposes. Master's thesis, School of Natural Sciences, Technology and Environmental Studies (2019)
18. Levac, D., Colquhoun, H., O'Brien, K.K.: Scoping studies: advancing the methodology. Implement. Sci. **5**(1), 1–9 (2010)

19. Li, W., Huang, H., Solomon, T., Esmaeili, B., Yu, L.F.: Synthesizing personalized construction safety training scenarios for VR training. IEEE Trans. Vis. Comput. Graph. **28**(5), 1993–2002 (2022)
20. Luiro, E., Hannula, P., Launne, E., Mustonen, S., Westerlund, T., Häkkilä, J.: Exploring local history and cultural heritage through a mobile game. In: Proceedings of the 18th International Conference on Mobile and Ubiquitous Multimedia, pp. 1–4. ACM (2019)
21. Luo, Y., Du, H.: Learning with desktop virtual reality: changes and interrelationship of self-efficacy, goal orientation, technology acceptance and learning behavior. Smart Learn. Environ. **9**(1), 1–22 (2022)
22. Manovich, L.: The Language of New Media. MIT Press (2002)
23. Metzger, S.A., Harris, L.M.: The Wiley International Handbook of History Teaching and Learning. Wiley, Hoboken (2018)
24. Moseikina, M., Toktamysov, S., Danshina, S.: Modern technologies and gamification in historical education. Simul. Gaming **53**(2), 135–156 (2022)
25. Parong, J., Mayer, R.E.: Learning about history in immersive virtual reality: does immersion facilitate learning? Educ. Tech. Res. Dev. **69**(3), 1433–1451 (2021)
26. Ramansyah, W., Aini, N., Fitriansyah, W., Pratama, M.D.: Virtual reality and educational game to learn Madurese history and alphabet for elementary school students. J. Phys. Conf. Ser. **1842**(1), 012012 (2021)
27. Remolar, I., Rebollo, C., Fernández-Moyano, J.A.: Learning history using virtual and augmented reality. Computers **10**(11), 146 (2021)
28. Sanderink, C.: History class in virtual reality for elementary education. Bachelor's thesis, University of Twente (2018)
29. Staley, D.J.: The pedagogical potential of virtual reality historical simulations. Simul. Gaming **44**(4), 559–568 (2013)
30. Tetep, R.M., Aziz, M.A., Husain, W.: Virtual reality application in education. J. Phys. Conf. Ser. **1242**(1), 012089 (2019)
31. Tricco, A.C., et al.: Prisma extension for scoping reviews (PRISMA-ScR): checklist and explanation. Ann. Intern. Med. **169**(7), 467–473 (2018)
32. Villena-Taranilla, R., Tirado-Olivares, S., Cózar-Gutiérrez, R., González-Calero, J.A.: Effects of virtual reality on learning outcomes in K-6 education: a meta-analysis. Educ. Res. Rev. **35**, 100434 (2022)
33. Villena Taranilla, R., Cózar-Gutiérrez, R., González-Calero, J.A., López Cirugeda, I.: Strolling through a city of the roman empire: an analysis of the potential of virtual reality to teach history in primary education. Interact. Learn. Environ. **30**(4), 608–618 (2022)
34. Xu, S.: Recommendation of online learning resources for personalized fragmented learning based on mobile devices. Int. J. Emerg. Technol. Learn. (iJET) **17**(3), 34–49 (2022)
35. Yildirim, G., Elban, M., Yildirim, S.: Analysis of use of virtual reality technologies in history education: a case study. Asian J. Educ. Train. **4**(2), 62–69 (2018)
36. Zainal, N., Awang, M.M., Ahmad, A., Ahmad, A., Muhthar, A.H.: A comparative study of history education curriculum in Malaysia and Indonesia. In: The 2nd International Conference on Sustainable Development and Multi-ethnic Society, pp. 86–89. Redwhite Press (2019)

Learning Content Management

Gradual Study Advising with Course Knowledge Graphs

Junnan Dong, Wentao Li, Yaowei Wang, Qing Li, George Baciu, Jiannong Cao, Xiao Huang(✉), Richard Chen Li, and Peter H. F. Ng

Department of Computing, The Hong Kong Polytechnic University, Hong Kong, People's Republic of China
{hanson.dong,wentao0406.li}@connect.polyu.hk,
{yaowei.wang,qing-prof.li,csgeorge,jiannong.cao,xiao.huang,
richard-chen.li,peter.nhf}@polyu.edu.hk

Abstract. Knowledge graphs (KGs) have been actively studied for pedagogical purposes. To depict the rich but latent relations among different concepts in the course textbook, increasing efforts have been proposed to construct course KGs for university students. However, the application of course KGs for real study scenarios and career development remains unexplored and nontrivial. First, it is hard to enable personalized viewing and advising. Within the intricate university curricula, instructors aim to assist students in developing a personalized course selection pathway, which cannot be fulfilled by isolated course KGs. Second, locating concepts that are important to individuals poses challenges to students. Real-world course KGs may contain hundreds of concepts connected by hierarchical relations, e.g., *contain_subtopic*, making it challenging to capture the key points. To tackle these challenges, in this paper, we present `GSA`, a novel gradual study advising system based on course knowledge graphs, to facilitate both intra-course study and inter-course development for students significantly. Specifically, (*i*) we establish an interactive web system for both instructors to construct and manipulate course KGs, and students to view and interact. (*ii*) Concept-level advising is designed to visualize the centrality of a course KG based on various metrics. We also propose a tailored algorithm to suggest the learning path based on what concepts students have learned.(*iii*) Course-level advising is instantiated with a course network. This indicates the prerequisite relation among different levels of courses, corresponding to the annually increasing curricular design and forming different major streams. Extensive illustrations show the effectiveness of our system.

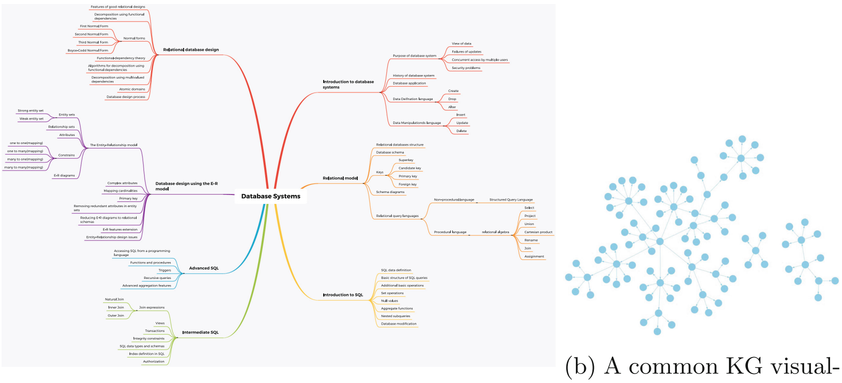
Keywords: Study Advising · Knowledge Graphs · Graph Visualization

1 Introduction

Knowledge graphs (KGs) have emerged as a potent tool for enhancing pedagogical achievements [9]. They could effectively represent the unstructured knowledge from textbooks as triples [6], i.e., (*Relational query languages*, contain_subtopic,

H. Xie et al. (Eds.): ICWL 2023, LNCS 14409, pp. 125–138, 2023.
https://doi.org/10.1007/978-981-99-8385-8_10

Procedural language), unveiling the latent connections between complex concepts and facilitating students' understanding rather than directly learning with the whole textbook. This particularly holds true and inspiring within the realm of college education. Recently, the course KG construction has gained much research attention [7]. Early studies propose deep learning methods to automatically construct KGs for education for their potential to illuminate complex relations among course concepts [2]. Another group of methods leverages online platforms [3,13], e.g., Wikipedia and MOOC, to enhance the concept extraction performance by linking the entities in the textbook with additional knowledge sources. However, though substantial efforts have been invested in constructing course KGs to capture the nature of concepts in course textbooks, they primarily focus on the construction and visualization of course KGs but fail to adequately address the practical difficulties encountered during students' utilization. A critical gap persists in bridging the course KGs with real-world study scenarios and subsequent career development. As educators seek innovative ways to empower university students with not only a good command of the course itself but also forming their own major stream through course selection, existing methods are not applicable to the downstream situation, also the related research remains limited and unexplored.



(a) A mind map format.

(b) A common KG visualization style via Neo4j.

Fig. 1. A real-world example of the course knowledge graph constructed for 'Database Systems'. It contains 96 triples connected by *contain_subtopic*.

Motivated by this, we investigate a tailored study advising system based on course knowledge graphs for both *intra-course* learning and *inter-course* career trajectories. However, this task is challenging for two major challenges.

First, isolated course KGs can hardly enable personalized viewing and advising. Within the intricate university curricula, instructors aim to assist students in developing individualized course selection pathways. To be associated with the aim of *inter-course* advising, isolated course KGs should be connected with each other in a logical way. An intuitive solution is shown in Fig. 2, which directly

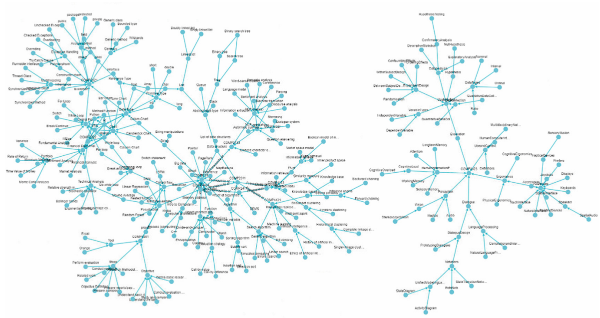


Fig. 2. An intuitive solution for students to obtain the overall academic picture by directly combining isolated course KGs.

combines all the course KGs in one, it is even harder to be interpreted. This makes study advising with course KGs for personalized course selection, as well as broader career planning, a nontrivial task. *Second*, it is inefficient for students to locate concepts that are important. In real-world scenarios, course KGs may contain hundreds of concepts connected by hierarchical relations, e.g., *contain_subtopic*, identifying crucial concepts within the complicated graph makes course KG-based study advising non-trivial. In Fig. 1, we visualize a course KG constructed for 'Database Systems', which is one of the smallest course KGs existing with merely 96 triples, in the form of a mind map for a clear illustration 1a. Despite the conciseness of this mind map, it is obviously time-consuming to read all the concepts for *intra-course* learning. The situations are significantly more complicated within the commonly used knowledge graph visualization with Neo4j database in Fig. 1b.

To this end, we present a novel approach, i.e., Gradual Study Advising (GSA), which leverages course KGs to effectively facilitate both intra-course comprehension and inter-course career development for university students. Specifically, (*i*) we first establish a basic interactive web-based system for university instructors to create and manipulate the suitable course KGs, granting them the tools to construct dynamic and informative graphs; (*ii*), we design concept-level advising, a novel visualization mechanism, corresponding to the *intra-course learning*, that quantifies the centrality of concepts within course KGs using a spectrum of metrics. Moreover, a substantial algorithm is proposed to tailor learning pathways based on students' acquired knowledge; (*iii*) A course-level advising is empowered by constructing a course network that uncovers the prerequisite relations among courses, enabling students to chart distinct career streams. This is inspired by the computer science curricula development in the Department of Computing, The Hong Kong Polytechnic University. For one CS freshman, four years are systematically designed from fundamental (year 1), broadening (year 2), and strengthening (year 3) to Specialization (year 4).

In general, our contributions are summarized below.

- We propose a new paradigm for gradual study advising with course knowledge graphs for intra-course learning and inter-course development.
- An interactive system is developed to facilitate both instructors to convey course knowledge and students to utilize it.
- Sufficient illustrations are provided to illustrate the effectiveness of our proposed system.

2 Functional Foundations for Interactive Web System

In this section, we introduce how we preliminarily prepare an interactive system for instructors to manipulate and publish a course KG to students, as well as for students to seek study advising. We first elaborate on the initialization of course KG construction based on textbooks in Sect. 2.1 and the tailored online manipulator for instructors in Sect. 2.2.

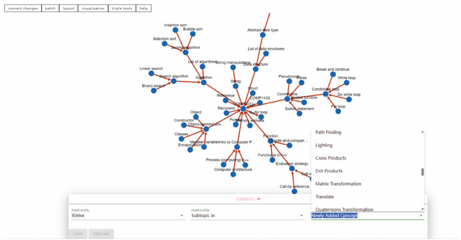
2.1 Course Knowledge Graph Construction

We undertake a comprehensive revision of the ontology of course KGs, which has been specifically tailored to better accommodate educational purposes in [7]. Our primary focus is the construction of 'contain_subtopic', which provides a clear picture of conceptual relationships hierarchically, thereby facilitating students' intr-course understanding and the establishment of links between different concepts. For each course, we began with a list of'seed entities', which are the core concepts identified by experienced educators that underpin the respective curriculum. Primary sources such as textbooks and Wikipedia were harnessed to build the course KGs, with a distinct emphasis on enlarging the graph centered around these seed entities. Then, we use the relation extraction model to output new triples, the input is a section of the textbook relevant to the seed entities parsed from a PDF file. For the extracted triples, we employed the seed entities for further filtering as we score the importance of each triple, ensuring that each EKG encapsulates the most essential information, thereby minimizing redundancy.

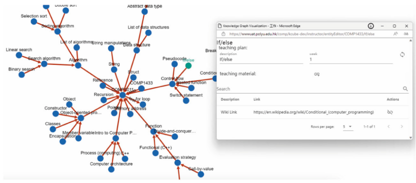
In particular, for computer science education in year 1: we identified approximately 20 distinct courses for our course KGs, each containing 10 to 20 seed entities. After the extraction, each course includes between 50 to 100 triples and an equivalent number of entities. Furthermore, we proceeded to merge and organize the different subgraphs obtained for each subject, which mainly involved entity alignment and redundancy checks. This step was primarily accomplished through natural language processing algorithms and manual rules.

2.2 Course Knowledge Graph Management for Instructors

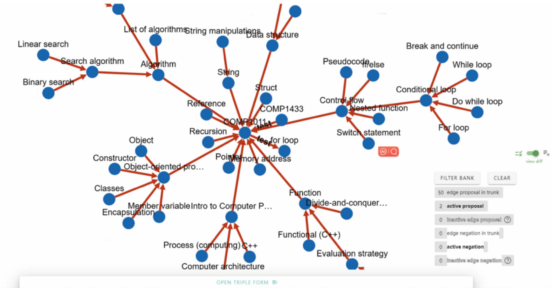
As a content management module for course KGs, in this section, we aim at providing management tools that support instructors' common activities in maintaining a course KG. Several managerial services are enabled by the included visualization interface and version control system. The visualization interface provides an orientation to the content and the relationship between them to users. The version control system expedites multiple managerial tasks that concern access right control, collaboration, and communication.



(a) A workspace for course KG manipulation.



(b) Managerial tools for course materials and links.



(c) Highlights of unstaged modification in the workspace. Unstaged triple creations are represented by solid lines. Unstaged triple removals are represented by solid broken lines. All staged triples are suppressed and turned semi-transparent.

Fig. 3. A simple illustration of the functionalities of the course KG manipulator.

Course Knowledge Graph Manipulator. Given the rapid changes in teaching targets based on students' feedback, to fulfill a convenient manipulation of course KGs, an integrated workspace is developed for instructors to easily manipulate course KGs and course material. Following the prevailing work [1], `GSA` provides a workspace interface (see Fig. 3), where instructors can manipulate the knowledge graph with the elementary operation, e.g., add or delete. For creation, the relation in a triple is defaulted as 'contain_subtopic'. Advanced operations, e.g. removal of nodes that are unreachable from the course node, are also available. In our manipulator, concepts are associated with the triples, each time an addition should take effect as a head/tail node of a triple, depicted in Fig. 3. Specifically, for new concept addition, instructors first input the name of a new concept, i.e., 'Newly Added Concept' and click the corresponding tail entity and click 'LINK' to generate triples in the graph. While for a new connection between existing concepts, manipulations can be easily done by either clicking the existing concept or choosing from the list.

While for managing the auxiliary learning materials, course material of an arbitrary concept node could be added through a pop-up window (see Fig. 3b). Information, such as providers, textual description, and the URL of the material, e.g., from Wikipedia, is stored by GSA, which also facilitates intra-concept learning.

Additionally, the workspace interface takes version control into consideration. The unstaged modification would not be tracked by the version control system. They are considered undocumented and volatile. Unstaged modification could be highlighted in an edit mode (see Fig. 3c). A workspace with unstaged modification could be set visible to students. This might be convenient in situations where the instructor has to expeditiously publish changes of the syllabus without the time to make precise remarks and tags for version control purposes. The downside of such an experience is that it may discourage instructors to proceed with a formal versioning process. As a countermeasure, highlighting unstaged remarks facilitates users when they want to track unstaged changes and reminds users to perform proper versioning and documentation tasks.

Algorithm 1. Algorithm of cross-tab synchronization using local Storage

```
Tab 1 for Instructor A:
Set the data to be shared using local storage.
End Tab 1

Tab 2 for Instructor B:
while listening for changes in local storage do
    If a change event occurs:
    if the changed key is 'sharedData' then
        Retrieve the new value from local storage.
        Display the updated value with Tab2: [received data].
    end if
end while
End Tab 2
```

Cross-Tab Synchronization. In consideration of the collaborative manipulation scenario among different instructors that may be responsible for the same course, they may work together in the workspace to create, edit, and manage course content. With cross-tab synchronization, changes made by one instructor are immediately reflected in all open tabs or instances where the course knowledge graph is being viewed or edited. This real-time collaboration ensures that instructors can see each other's changes without delays, fostering efficient teamwork.

We explain this process in Algorithm 1. Specifically, in the instructor A's tab, while making changes to the course KG, the instructor updates the graph according to their actions. After each update, a synchronization event is triggered to notify other tabs that changes have been made. In other tabs where the same course knowledge graph is being viewed or edited, a continuous listening loop monitors synchronization events. Upon detecting a synchronization event, these tabs receive the updated course knowledge graph data and update their displayed knowledge graphs to mirror the changes made by the instructor.

This process ensures that all tabs displaying the course KG remain in sync, providing instructors with a cohesive and real-time collaboration environment. The cross-tab synchronization approach for instructors' manipulation of course KGs offers several benefits:

- Real-time Collaboration: Instructors can collaboratively work on the same course content, seeing each other's changes in real-time.
- Seamless Experience: Changes made by one instructor are immediately reflected across all tabs, eliminating confusion or discrepancies.
- Enhanced Productivity: Instructors can focus on content creation and manipulation without interruption or manual updates.

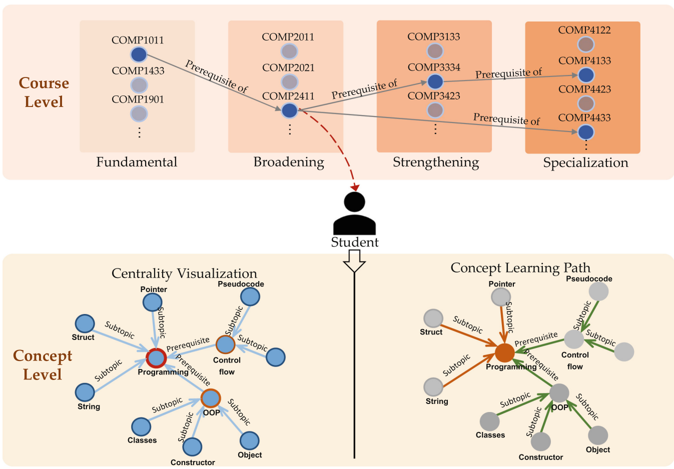


Fig. 4. A dual framework for study advising from course level and concept level.

3 Approach: GSA

As illustrated in Fig. 4, we propose a gradual study advising framework that consistently integrates both intra-course learning and inter-course advising. We aim to first guide the students to form their personalized major stream, as well as the career development through inter-course advising, and then dig deeply into particular courses with the intra-course learning module.

3.1 Inter-course Advising

Course Network Construction. In order to provide gradual inter-course advising, we have developed a sophisticated course network that interconnects all the courses through the "prerequisites of" relationship. Instead of combining all the course KGs together, m this innovative approach significantly reduces the graph size and highlights the progressive course-level relations.

We draw inspiration from the curricular design framework utilized by the Department of Computing at The Hong Kong Polytechnic University. Following a meticulously structured pathway, CS freshmen undergo a four-year journey, progressing systematically from fundamental (year 1) to broadening (year 2), strengthening (year 3), and culminating in specialization (year 4).

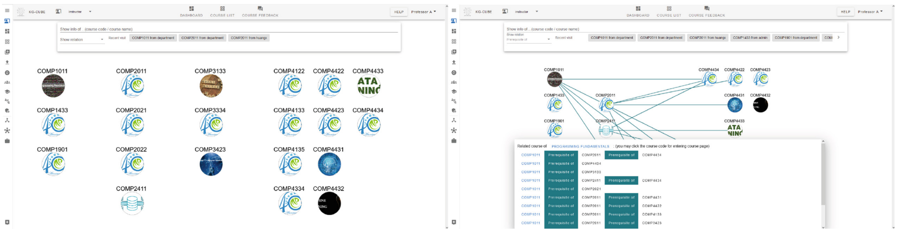


Fig. 5. The intra-course advising provides visualization of prerequisite relations among courses. Students can toggle the visualization. `GSA` also allows them to open a list that includes all textual paths that contain particular courses.

Relational Path Finder. Following this transformative approach, students are empowered to navigate the course network according to their year level, aligning their course selections with their career aspirations and major streams. For instance, in the Department of Computing at the Hong Kong Polytechnic University, those who aspire to become fin-tech experts can strategically choose courses in a coherent sequence. In year 1, they could select fundamental knowledge, easing their transition to university studies with an introduction to Scheme. In the second year, they could select the courses, of which they have taken the prerequisites, to acquire broad computing skills, along with rudimentary concepts of economics, accounting, and finance. The third year is dedicated to continuing to strengthen core competencies, encompassing software engineering, systems security, and a selection of computing or finance electives. Finally, in the fourth and final year, students specialize in areas like artificial intelligence, machine learning, and pattern recognition, as well as emerging fields like crowdfunding, e-finance, and e-payment systems. This curriculum-guided inter-course advising ensures a logical progression, enabling students to make informed and strategic choices in line with their evolving career goals.

To achieve this goal, we design a relational path finder where prerequisite relations between courses could be visualized through `GSA`. Visualization of these relationships is available in the homepage. Particularly, the visualization of the prerequisite relationship is illustrated in Fig. 5. Such visualization orients users about the overall structure of the university program. For example, the prerequisite relationship among courses in this case could help students to decide what course they should select or review for future stream development.

3.2 Intra-course Advising

To tackle the challenges that course KGs are difficult to be utilized by students given the number of concepts and complicated connections, in this subsection, we employ different centrality metrics such as degree centrality, and PageRank, to clearly depict the centrality of one course KG, providing students with a dynamic visual representation toolbar that could highlight the significance of concepts within the course KG. These visualizations enable students to quickly grasp the core ideas and critical nodes within the knowledge graph, promoting efficient learning. The centrality metrics illuminate nodes with high connectivity, bridge nodes that connect disparate areas, and influential nodes that carry substantial importance. Through interactive and intuitive visualizations, students can identify pivotal concepts, explore relationships, and navigate the course KG's complexity with ease. These methodologies not only facilitate the rapid acquisition of key knowledge but also empower students to comprehend the interconnections that underlie the course content, fostering a deeper understanding of the subject matter.

In this paper, we showcase two aspects of centrality visualization by employing 'Degree' and 'PageRank'.

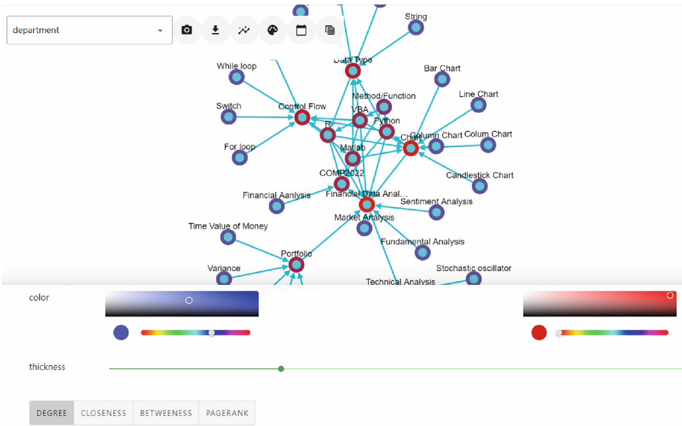


Fig. 6. The centrality visualization based on degrees.

Degree Centrality. In the realm of intra-course learning, the visualization of centrality through the lens of degree centrality emerges as a powerful tool. The concept of degree centrality brings forth a structured approach to understanding the pivotal nodes within a course's knowledge graph. This visualization technique is underpinned by the calculation of the degree of a concept, which reflects its connectedness to other concepts within the graph. In essence, the degree centrality $deg(e_i)$ of a concept e_i, encapsulates the sheer number of relationships linked to it. Mathematically articulated as:

$$deg(e_i) = Len_{r \in N_{e_i}}(r), \tag{1}$$

where N_{e_i} is the one-hop neighbor triples centered by concept e_i, and r represents the relation appears in N_{e_i}. This metric holds profound significance. By quantifying the number of edges or relationships incident to an entity, degree centrality provides a quantitative representation of its influence and importance within the graph. In Fig. 6, we visualize the central concepts with colors from shallow to deep.

In the context of intra-learning with course KGs, this visualization approach serves as a compass, guiding learners toward the concepts that play a pivotal role in shaping their understanding of key concepts. Through degree centrality visualization, students gain an intuitive grasp of the central components that underpin the course's knowledge structure, enhancing their ability to navigate and comprehend complex subject matter.

PageRank Centrality. Similar to degree centrality, we visualize the result of PageRank in Fig. 7.

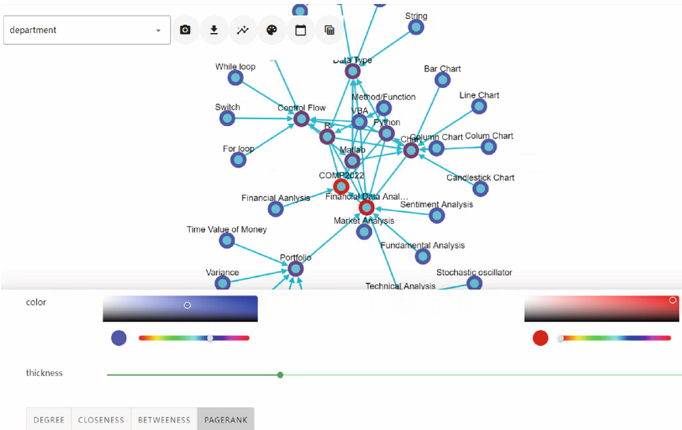


Fig. 7. An alternative visualization of centrality based on the 'PageRank' metric.

The adapted PageRank equation for a knowledge graph, considering entities and their relationships, is formulated as follows:

$$PR(E_i) = (1 - d) + d \times \sum_{e_j \in N_{e_i}} \frac{PR(e_j)}{|L(e_j)|} \tag{2}$$

where $PR(E_i)$ is the PageRank score of concept e_i. $e_j \in N_{e_i}$ is the set of neighbor concepts that link to entity e_j. d is the damping factor, a value between 0 and 1, representing the probability that a student follows a connection rather than jumping to another random concept.

Comparing 'Degree' and 'PageRank' as centrality metrics, we adopt distinct approaches to visualize the importance of nodes from different views. Degree centrality focuses on the straightforward notion of connectedness. By counting the number of relationships linked to an entity, 'Degree' identifies nodes with high interaction and participation within the graph. Intuitively it is computationally simple for rendering and quickly identifying heavily connected concepts. However, it might overlook concepts that are indirectly influential due to their position in the graph. While 'PageRank' introduces a more nuanced perspective by considering not only the number of relations but also the quality of those connections. This metric reflects the importance of one concept which is not solely determined by its own degree but also by the importance of concepts linking to it. This algorithmic approach accounts for the graph's structure and provides a more sophisticated understanding of influence.

In general, for intra-course learning in our `GSA`, different choices of centrality depend on the learning objectives and the nature of the course KGs. We provide a balanced approach that might involve employing various metrics, leveraging all their insights for a more refined understanding of centrality and influence.

3.3 Concept Learning Path Recommendation

In addition to the centrality visualization, for advanced intr-course learning with course KGs, we also propose a tailored recommendation algorithm that still remains unimplemented. As shown in Fig. 4, given a course KG, denoted as C, we would like to do personalized and time-sensitive recommendations based on the semester teaching schedule and how well the student grasps the current progress. We design an expectation score that evaluates the importance of one concept e_i that should be recommended for the student s to preview or review based on the schedule. The proposed recommendation is formulated as a ranking problem. `GSA` will first calculate the expectation score $E(e_i, s)$ with a multiplication of: (i) the importance score of concept e_i, $deg(e_i)$; (ii) the relatedness of the concept R_{N_i} considers how many of the subtopics/prerequisite concepts have been taught according to current progress, the equations are derived as:

$$E(e_i, s) = deg(e_i) \times R_{N_i} \tag{3}$$

Within each interaction, `GSA` will first traverse the academic calendar and the semester schedule for particular courses. Then a personalized study progress will be retrieved based on the timestamp which indicates the courses and the concepts in each course that the student has taken. Finally, we calculate the expectation score for each candidate concept and orderly sort them by rankings. Those with higher scores are expected to be recommended to students for their review/preview subject to the temporal state.

4 Related Work

4.1 Educational Knowledge Graph Construction

Knowledge graphs have gained traction in education for enhancing the representation and navigation of educational content. In early studies, KnowEdu [2] proposes to combine deep learning and rule mining methods, i.e., GRU and p-Apriori, to extract knowledge from internal system data and evaluate it with two human experts' annotation on all entities and relations. Recently, efforts have been made to leverage various e-learning platforms, e.g., Wikipedia and MOOC. EduKG [13] constructs knowledge graphs from educational resources to aid content recommendation. MOOCKG [3] links course concepts with external knowledge sources to enrich course content. In this paper, we bridge the gap between educational KG construction and effective academic advising for students with our GSA system.

4.2 Study Advising

The field of study advising has witnessed substantial advancements in recent years [10]. Several notable studies and systems have contributed to the understanding and implementation of study advising [1,4,5]. Early advisors have typically relied on face-to-face interactions to provide guidance to students. These interactions often involve discussions about course selection, career paths, and academic progress. Garton et al. [11] emphasize the importance of personalized interactions in their study on student perceptions of academic advising. Sweker et al. [12] investigate the importance of the number of meetings between advisors and first-generation students. Recently, with the advancement of technology, various digital tools have been developed to enhance study advising processes. Online platforms, such as advising portals and educational planning software, have been designed to facilitate communication between students and advisors. MacDonald et al. [8] suggest the power of distant advising through online platforms, revealing the impact of technology-enhanced advising platforms on student engagement and satisfaction. In our GSA, we empower instructors with an intelligent interactive system to manipulate the course knowledge graph and automatically provide both intra-course and inter-course advising.

5 Conclusions and Future Work

In this paper, we present a novel Gradual Study Advising system, i.e., GSA that emerges as a pioneering solution to bridge the gap between course knowledge graph (course KG) construction and students' practical needs of learning and career development with course KGs. The fundamental role of course KGs in reshaping pedagogical approaches has led us to formulate a tailored advising system that integrates both intra-course learning and inter-course career. Specifically, we empower university instructors with an interactive web-based platform that allows them to craft dynamic and informative course KGs, ensuring the quality of comprehensive course KGs. Moreover, our innovative concept-level advising mechanism transforms course KGs into easily understandable visualizations, by quantifying concept centrality through a range of metrics. This enables students to comprehend the hierarchy of concepts within course KGs and paves the way for personalized learning pathways. Finally, we extend our approach to inter-course advising, establishing a course network that uncovers prerequisite relations among courses, and guiding students to chart distinct career paths inspired by the progressive curricula design in the Department of Computing at the Hong Kong Polytechnic University.

In future work, we will continue implementing the auto-advising on concept learning path recommendation based on personalized study achievements. By incorporating adaptive learning algorithms, GSA will dynamically tailor learning pathways and career trajectories based on individual student preferences and their taken concepts, as well as courses, according to the semester schedule. This would amplify the effectiveness of our system in guiding students toward their academic and professional goals.

Acknowledgement. This work was supported in full by the Hong Kong Polytechnic University, Project of Strategic Importance (project number: P0036846).

References

1. Chan, J.C.K., et al.: Intelligent instructional design via interactive knowledge graph editing. In: González-González, C.S., et al. Learning Technologies and Systems. ICWL SETE 2022 2022. LNCS, vol. 13869, pp. 41–52. Springer, Cham (2023). https://doi.org/10.1007/978-3-031-33023-0_4
2. Chen, P., Lu, Y., Zheng, V.W., Chen, X., Yang, B.: KnowEdu: a system to construct knowledge graph for education. IEEE Access **6**, 31553–31563 (2018)
3. Dang, F., Tang, J., Li, S.: MOOC-KG: A MOOC knowledge graph for cross-platform online learning resources. In: 2019 IEEE 9th International Conference on Electronics Information and Emergency Communication (ICEIEC), pp. 1–8. IEEE (2019)
4. Hutson, B.L.: The impact of an appreciative advising-based university studies course on college student first-year experience. JARHE **2**(1), 4–13 (2010)
5. Junco, R., Mastrodicasa, J.M., Aguiar, A.V., Longnecker, E.M., Rokkum, J.N.: Impact of technology-mediated communication on student evaluations of advising. NACADA **36**(2), 54–66 (2016)
6. Kejriwal, M.: Domain-Specific Knowledge Graph Construction. SCS, Springer, Cham (2019). https://doi.org/10.1007/978-3-030-12375-8
7. Li, W., et al.: Constructing low-redundant and high-accuracy knowledge graphs for education. In: González-González, C.S., et al. Learning Technologies and Systems. ICWL SETE 2022 2022. LNCS, vol. 13869, pp 148–160. Springer, Cham (2023). https://doi.org/10.1007/978-3-031-33023-0_13
8. MacDonald, M., Garcia, B.W.: Advising the online student: a breakout of advising frequency, preferences, and satisfaction of online students. IJDLA **23**(3), 13 (2020)
9. Rizun, M., et al.: Knowledge graph application in education: a literature review. Acta Universitatis Lodziensis. Folia Oeconomica **3**(342), 7–19 (2019)

10. Steele, G.E.: Student success: academic advising, student learning data, and technology. NDHE **2018**(184), 59–68 (2018)
11. Suvedi, M., Ghimire, R., Millenbah, K., Shrestha, K.: Undergraduate students' perceptions of academic advising. NACTA J. **59**(3), 227–233 (2015)
12. Swecker, H.K., Fifolt, M., Searby, L.: Academic advising and first-generation college students: a quantitative study on student retention. NACADA **33**(1), 46–53 (2013)
13. Zhao, B., et al.: EDUKG: a heterogeneous sustainable k-12 educational knowledge graph. arXiv preprint arXiv:2210.12228 (2022)

Intended Learning Outcomes and Taxonomy Mapping at University Level

Florian Eckkrammer[1](✉), Harald Wahl[1], and Luis Torres Pereira[2]

[1] University of Applied Sciences Technikum Wien, Vienna, Austria
{eckkramm,wahl}@technikum-wien.at
[2] University of Trás-os-Montes e Alto Douro, Vila Real, Portugal
tpereira@utad.pt

Abstract. At the University of Applied Sciences Technikum Wien, the intended learning outcomes (ILO) for individual study programs are well defined. These ILO are derived from qualification profiles and should ensure well-educated graduates for professional success. However, at the university level across several study programs, a lack of coordination in ILO development exists. A comparison across study programs and individual courses can show synergies of curricula. It can identify course similarities across programs, allowing collaborative development and standardization with the aim of cost-effective quality improvement. Thus, this paper proposes a solution to this challenge by harmonizing ILO and employing taxonomies for clear outcome classification. Therefore, text analysis, text enrichment with additional information, taxonomy mapping, and the annotation of the intended learning outcomes are the main steps of the prototype.

Keywords: Intended Learning Outcomes · Educational Taxonomies · Mapping of Taxonomies

1 Problem Description

Although a university operates several study programs, in terms of learning outcomes, a study program is the central point of reference. A qualification profile [1] is defined as part of the creation process of a curriculum. This profile should ensure that after completion of the study program the necessary skills for a successful assertion in the world of work are available. Derived from this are competencies, which describe acquired skills, knowledge and abilities and are described in the form of intended learning outcomes (ILO) [2, 3]. The curriculum is shaped through the placement of learning outcomes in courses and/or modules and the alignment of the courses/modules through the semester.

Over several years, a multitude of ILO are developed at universities within the scope of the multitude of study programs as well as their continuous development. This development is not coordinated beyond study program/course level. A harmonization of the ILO can be achieved in terms of content as well as in terms of formulation. The content-related alignment focuses on ability, knowledge, and skills. The formulation deals with

H. Xie et al. (Eds.): ICWL 2023, LNCS 14409, pp. 139–146, 2023.
https://doi.org/10.1007/978-981-99-8385-8_11

the structure of the description of the learning outcome. If ILO are formulated by using different taxonomies, comparison, and coordination between them could be difficult.

A holistic overview of learning outcomes at the university level does not exist in most cases. However, this would be necessary to enable a comparison across courses and to identify synergy effects in the development of courses. Furthermore, the overview could be used to make statements about the quality of the formulation of individual learning outcomes.

In order to distinguish from existing research activities around ILO, a classification of the activities into the following areas (or views) is made:

- Process view of/around ILO
- Formulation and grouping of ILO
- Monitoring in general
- Efficiency
- Social aspect

The process view describes on course level when and how learning outcomes are developed. The Constructive Alignment approach according to [3] is a prominent example.

The taxonomies, their grouping, delimitation, or arrangement internally and externally emerge from the research activities that are concerned with the formulation and grouping. (e.g. [7, 8]). Also, the (partly) hierarchical structure within a taxonomy is treated in these investigations.

Monitoring aspects are considered, for example, in the context of learning analytics projects and focus on the behavior of learners and teachers to draw conclusions about courses, materials, methods and the like.

Studies related to efficiency focus on the interaction between ILO the course (or learning material) as well as the students.

Studies on social aspects look at issues between students and ILO.

While there is more or less extensive research on the areas listed above, research examining ILO and their (content) relationships to each other at the organizational level is rare to nonexistent. Adjacent to this research are research activities around the graphical representation of competencies as well as learning maps and knowledge maps.

2 Structure of Intended Learning Outcomes

The Europe-wide harmonization of degrees in the Bologna Process, as well as the definition of the European Qualifications Framework (EQF) for better comparison of (degree) certificates, has led to a trend reversal in the mapping and development of curricula. Instead of content, qualifications are primarily described at the course or module level (whereby modules represent combinations of courses) based on ILO. This follows the recommendation of the European Union, which states that "an approach based on learning outcomes should be used to describe and define qualifications" [4].

Thus, there is a need for a qualitative learning outcome formulation that is as clear and consistent as possible. Furthermore, it is necessary to conclude from the set of learning outcomes to the overarching competency areas of a degree program [5].

According to [3], learning outcomes consist of:

- a level of understanding
- content
- context.

The verb describes the "level of understanding". The object to which the verb refers describes the content and the context describes the discipline to which the verb relates. Very often, adjectives relativize the content and/or the context. For example, a certain skill may refer to "simple programming tasks" rather than complex tasks. The distinction between the different qualifiers often remains a gray area and is not further specified.

To represent the level of understanding in a classified way, learning outcomes make use of taxonomies. The basis of the classification is very different for the different taxonomies. In the context of the solution of the described problem only Bloom's taxonomy [6, 7] were implemented in a first step. However, the solution approach offers the possibility to consider further taxonomies. In particular, the SOLO taxonomy [8, 9] can be easily added due to its similar structure. Mappings between the taxonomies could also be implemented in a further step.

3 Concept of Solution and Prototype

Figure 1 shows the processing chain of the intended learning outcome. After passing through the text analysis agent, the texts can be enriched with additional information. After this (optional) step, a mapping into the selected taxonomy is performed. The result is an annotated learning outcome that can be further processed as needed.

Fig. 1. Chain of intended learning outcome (ILO) workflow

3.1 Intended Learning Outcome

An intended learning outcome, which could currently be processed by the prototype could look like this:

"After passing this course successfully students are able to (plan and coordinate[1]) a (project[2]) in (small[3]) (groups[4])."

According to the above structure of ILO, the area marked in green (1) represents the verb and reflects the "level of understanding". The orange text (2) describes the content. The text marked in blue (4) indicates the context. The text marked in red (3) represents a limitation of the described competence. Text parts marked in black are ignored by the system.

3.2 Text Analysis Agent

The workflow was implemented in the system in such a way that different algorithms can be used for text analysis. Currently implemented algorithms are based on [5]. The result of the analysis workflow is a segmentation into subject, predicate and object. Adjectives are mapped according to Universal Dependencies [10].

3.3 Enrichment Agents

Several Agents could be used to enrich the preprocessed text. The currently available Agent uses APIs to query synonyms for the verb, the object and the context. This information is used later on to cluster the integrated learning outcomes.

3.4 Taxonomy Mapping

The enriched learning outcome is checked against existing taxonomies to contextualize the ILO to the appropriate level of the selected taxonomy. The current prototype supports

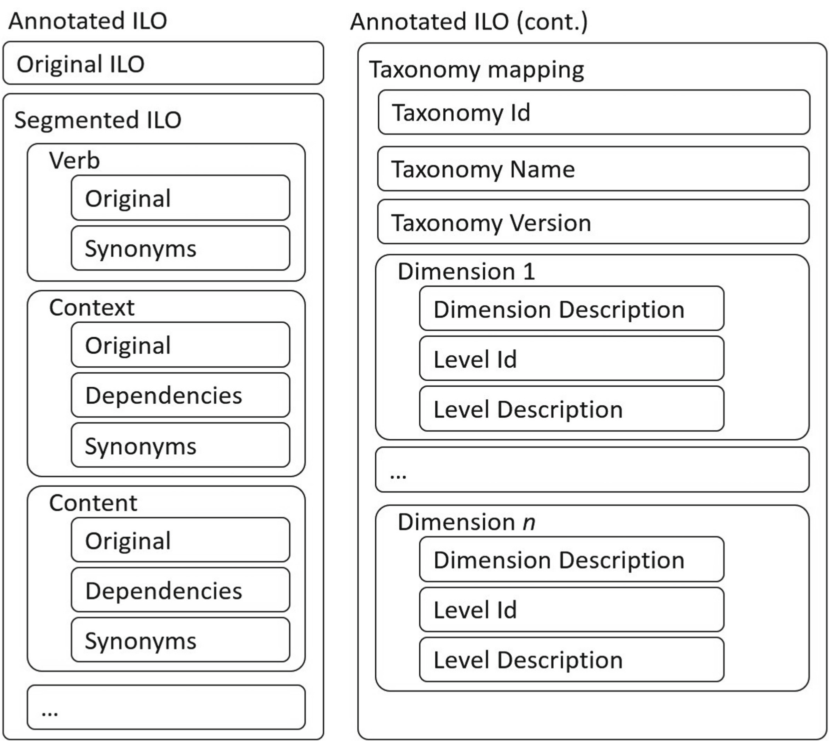


Fig. 2. Structure of annotated learning outcome formatted as json (for better presentation the sequential content is displayed in two columns)

the Blooms [6] taxonomy and the Anderson & Krathwohl [7] taxonomy. To improve the mapping to the different taxonomy levels synonyms were used and stored in a database[1].

3.5 Annotated Intended Learning Outcome

The annotated learning outcome is formatted in json format and contains the following sections as shown in Fig. 2.

To cover the requirements of the different taxonomies the json format supports multiple dimensions. The Blooms Taxonomy for example needs only one dimension. The SOLO taxonomy for example needs two dimensions.

4 Potential of Standardization

Universities in both North America and Europe are under substantial pressure.

The changing relationship between the state and higher education, cultural shifts, and broad trends toward globalization have led to financial pressures on universities and intensified competition among them. Universities have responded to these pressures by cutting costs, becoming more entrepreneurial, increasing administrative control, and expanding the use of rationalized tools for management [11].

One way to reduce costs while improving quality is to identify similar or largely identical courses that are offered in different programs. People who develop courses independently of each other can be grouped together and share the development costs. Or teams of course developers and delivery teams can be established. In particular, courses that are to be taught in parallel over several groups in the same year benefit from centralized development and rollout. Further development, however, is only carried out by one course, not by several courses in parallel. If similarities between courses are identified with the help of the described approach, it can be checked whether the courses can be aligned to one centralized course. A comparison of 12 bachelor programs of the UAS Technikum Wien with a total of 377 courses [12] showed a standardization potential of 43% that could be standardized across several faculties. Another 25% could be standardized within one faculty. A faculty in that case maintains between 2 and 4 bachelor programs. Only 32% of the courses were program specific. The average development time of a 5 ECTS course was 280 h. Measured according to the standardization potential, this resulted in savings in the context of a new development, considering new didactic concepts such as constructive alignment as planning approach and blended learning, of several full time equivalent (FTE). In the context of further development efforts of the course materials, this savings potential increases further. However, due to the high level of granularity of standardization (at the course or module level), differentiating features of courses may be reduced or lost entirely. To compensate this disadvantage, standardization on learning outcome could be done.

[1] https://www.teachthought.com/learning/what-is-blooms-taxonomy/

5 Further Steps

The existing prototype enables the delivery of annotated ILO and mappings for further processing purposes.

The next step is to represent the formulated learning outcomes of the entire university in the context of a competence map. Existing research on possible representations [13] can be taken up and implemented. This representation can support applicants in their choice of studies. Furthermore, the study programs can use such a representation to sharpen their profile and to distinguish themselves from other study programs.

Courses that are very similar in terms of learning outcomes can be developed together. In order to address small distinctions, it would be possible to create variants.

Currently, faculty staffing is strongly content-driven. If the constructive alignment approach is consistently developed further, it would also be feasible to link the ILO to a competence profile of lecturers. The assignment of lecturers to courses could then be competence-based in coordination with the ILO.

Within the framework of the Bologna Process, the European Union has determined that accreditations are to be carried out on the basis of existing competencies. The ILO-based representation could therefore support students and administrators to make the crediting processes more efficient.

Regarding the algorithms that are currently implemented, both the text analysis process and the mapping of the taxonomies can be enhanced. Supporting AI-based approaches would be easily possible due to the flexible software design.

The implementation of additional taxonomies is also envisioned. The multi-dimensional representation in json format already supports this development step. A mapping of the taxonomies would be possible, as long as the dimensions are comparable.

An essential development step is the component-oriented standardization of courses based on ILO. To avoid the problems of course-based standardization, ILO can be understood as building blocks of courses. A new course is created from a pool in the sense of a modular system by a competent development team. ILO that do not exist can be added to the pool.

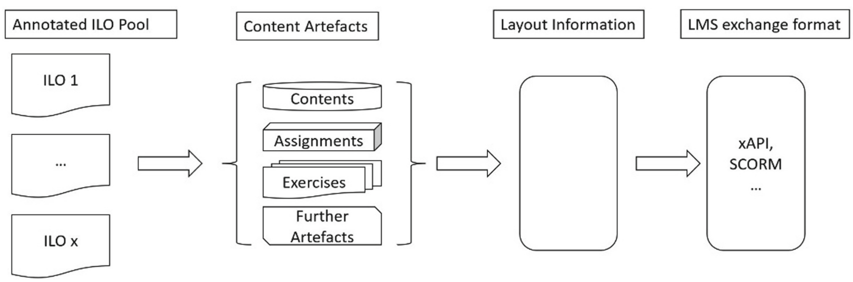


Fig. 3. Component based course design based on annotated learning outcomes to enable semi-automated course design

As shown in Fig. 3 the annotated learning outcomes have a connection to the learning artefacts which are necessary to achieve the designated learning outcome. These artefacts can be content, assignments, exercises and other artefacts. Combined with layout information the information can be structured by using existing LMS exchange format like xAPI or others.

6 Conclusion and Limitations

The described approach provides the basis for a wide range of applications and can be used in many ways, as described in the previous chapter, to represent interdependencies at the curricular level and subsequently to align ILO at the institutional level. A qualitative assessment of the formulation of ILO can be supported but not fully automated in the current state of development. Further development work will be required for this.

The approach cannot provide information on whether the learning materials or the course structure comply with the ILO and whether the learning outcomes can be achieved. Nor can it verify whether any constructive alignment planning approach has been implemented. Currently, only ILO are considered and statements about formulation, and their relationships to each other are made. Relationships to content artifacts, layout, course structure and examination structure are not targeted. Consequently, no judgement can be made as to whether an exam structure is suitable for checking the achievement of ILO. Also, the approach per se does not lead to standardization and alignment of curricula. Organizational measures are necessary for this which could be supported by additional information generated by the described approach.

References

1. Baldauf-Bergmann, K., Mischun, K., Müller, M.: Leitfaden zur Formulierung und Nutzung von Lernergebnissen. Von Netzwerk Studienqualität Brandenburg (2013)
2. Schermutzki, M.: Lernergebnisse - Begriffe, Zusammenhänge, Umsetzung und Erfolgsermittlung. In: Handbuch Qualität in Studium und Lehre. Evaluation nutzen – Akkreditierung sichern – Profil schärfen. Stuttgart, Berlin (2007)
3. Biggs, J., Tang, C.: Teaching for Quality Learning at University, 4th edn. Open University Press, London (2011)
4. European Union: Recommendation of the European Parliament and of the Council of 23 April 2008 on the establishment of the European Qualifications Framework for lifelong learning. European Union (2008)
5. Karpenko, A.: Automated extraction and matching of information from learning outcomes texts, Vienna (2017)
6. David, R.: A revision of bloom's taxonomy: an overview. In: Theory into Practice, vol. 41, no 4 (2002)
7. Anderson, L., et al.: A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives. Longman, New York (2001)
8. Lucander, H., Bondemark, L., Brown, G., Knutsson, K.: The structure of observed learning outcome (SOLO) taxonomy: a model to promote dental students' learning. Eur. J. Dental Educ. Off. J. Assoc. Dental Educ. Eur. **14**, 145–50. https://doi.org/10.1111/j.1600-0579.2009.00607.x

9. Huilan, S., et al.: Educational management in Critical Thinking Training Based on Bloom's Taxonomy and SOLO Taxonomy. In: International Conference on Information Science and Education (ICISE-IE), Sanya, China (2020)
10. Kondratyuk, D., Straka, M.: 75 languages, 1 model: parsing universal dependencies universally. In: Conference on Empirical Methods in Natural Language Processing, Hong Kong (2019)
11. Berman, E., Paradeise, C.: Introduction: the university under pressure. In: The University Under Pressure (Research in the Sociology of Organizations), vol. 46, pp. 1–22. Emerald Group Publishing Limited, Bingley (2016)
12. Lünenborg, Y.R., Lanickova, K., Preusche, I.: Implementierung eines didaktischen Konzepts in sämtliche Bachelor-Lehrveranstaltungen der FHTW für mehr Lehr-/Lernqualität: Tübingen (2022)
13. Suwignyo, H., Kurniawan, C., Kuswandi, D., Karkono, Ariani D.: Robust data mining visualization for learning outcomes. In: 2nd International Conference on Information Technology and Education (ICIT&E), Malang, Indonesia (2022)

Mixed Reality Learning Visualizations Using Knowledge Graphs

Benedikt Hensen(✉), Alexander Rechtmann, and Alexander Tobias Neumann

Chair of Computer Science 5, RWTH Aachen University, Aachen, Germany
{hensen,neumann}@dbis.rwth-aachen.de

Abstract. Learners have limited methods such as mind maps to gain an overview of the structure and key facts in their learning material. Knowledge Graphs (KGs) are a more powerful way of representing information on a learning topic since computers can also work with encoded knowledge and can help populate the graph with facts. However, visualizing the content of a KG to make it understandable and applicable for learning is a challenging task. We investigated the effect that different node representations of 3D Mixed Reality (MR) visualizations generated from KG data have on the learning behavior, memorability and user acceptance of such novel educational applications. For this, we developed an open-source software artifact that converts KGs about learning materials into three-dimensional networks which students can explore in a MR environment. The results of the study show that users prefer node visualizations in the graph that convey a visual identity in the form of images as opposed to merely displaying nodes as spheres. This preference is supported by the measured memorization, usability and technology acceptance values during the study. These insights help find an effective visualization of KGs for learning and understanding systematic information with their semantic interconnections in MR.

Keywords: Mixed Reality · Knowledge Graphs · Technology-Enhanced Learning · Graph Visualization

1 Introduction

With the digitalization of educational material, there is a large potential to store the learning content in a semantic way in KGs. KGs are a semantic way of representing information by modeling the relationships between facts. A common way of storing KGs is Resource Description Framework (RDF) where facts are expressed as triples which consist of a subject, a predicate and an object [11].

The research leading to these results has received funding from the German Federal Ministry of Education and Research (BMBF) through the project "Personalisierte Kompetenzentwicklung und hybrides KI-Mentoring" (tech4compKI) (grant no. 16DHB2213).

H. Xie et al. (Eds.): ICWL 2023, LNCS 14409, pp. 147–154, 2023.
https://doi.org/10.1007/978-981-99-8385-8_12

Hence, the subject and object form the nodes of the graph and the triples form connections between them via the predicates. The advantage of using KGs is that they are machine-readable which means that expert systems can be built on top of the data source to quickly retrieve knowledge or to reason based on the existing information. So, by encoding learning materials as KGs computers can support the students, e.g., by providing facts from the knowledge when required and can use the given information to validate of student answers. The contained information can also provide an overview of the learning content when visualized. However, understanding the structure of a KG and memorizing the information contained within can be challenging, especially when using traditional 2D visualizations of such graphs. In 2D, there are intersections between edges in the graph which lead to visually ambiguous paths and visual clutter. This can be avoided in 3D by introducing the depth of the node as an additional parameter. Bringing the 3D graph visualization into a MR environment tends to the student's spatial memory and provides orientation. In MR, the graph can be anchored to locations in the real world. In this paper, we investigate the research question which impact different node representations have on the learning behavior of students using KG visualizations in MR. We implemented a Technology Enhanced Learning (TEL) system in MR which can visualize the content of KGs in MR. Our contribution is a novel approach to represent the resulting graph's nodes as 3D models or images to improve the memorability in the learning process. The paper is structured as follows: In Sect. 2, we examine similar research approaches. Section 3 covers the applied methodology. The implementation is evaluated in Sect. 4. We conclude the paper with Sect. 5.

2 Related Work

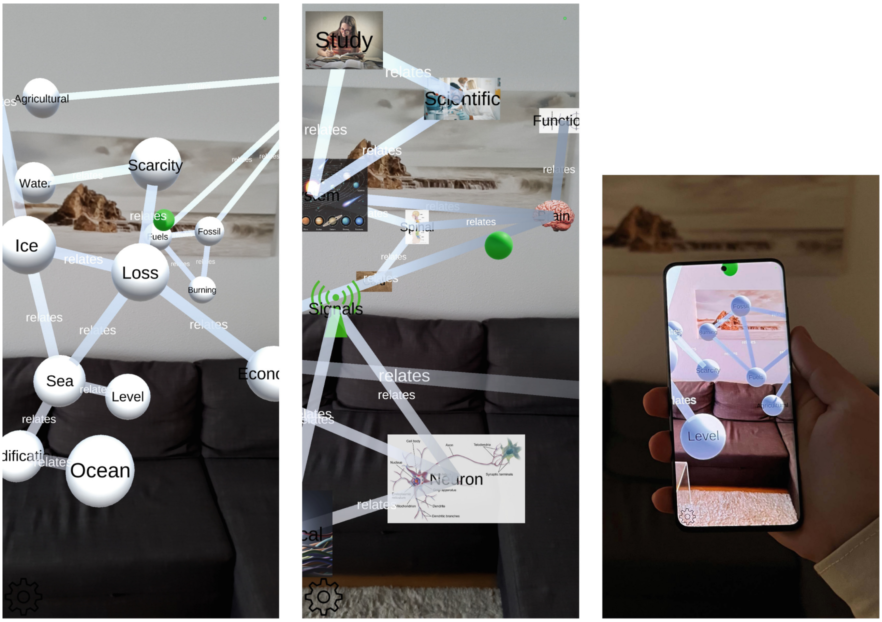
The related work contains similar approaches for investigating 3D visualizations of general graphs and of KGs. An overview of 3D layout options is given by Herman et al. [7]. In their study, they show that the 3D environment offers more space for the layout and is intuitive to navigate. A challenge of these 3D visualizations concern occlusion and disorientation. However, they conclude that Virtual Reality (VR) technology can mitigate these factors. Whereas our approach inspects the node representations, Büschel et al. [2] looked at different edge representations in Augmented Reality (AR). They investigated directed and undirected edges and visualized them using, e.g., straight lines, dashes, curves and sloped connections which indicate direction. A previous 3D visualization of graphs has been implemented by Drogemuller et al. [4]. Their system VRige takes social network data as input and visualizes them for the HTC Vive. Users can interact with the nodes and immersive 3D interaction elements to explore the graph with filters. Another VR graph visualization was realized by Kortemeyer [8]. The system uses a force-directed layout based on the Fruchterman-Reingold algorithm [5]. This allows the graph to be positioned dynamically and to react to user inputs such as dragging nodes to different locations. The Remote Sensing Scene Knowledge Graphs by Potnis et al. [10] makes use of KGs and the spatial positioning abilities. They encode data of flood events in a knowledge graph and position the data nodes of the graph on a map. The information are directly connected to the location that the data is expressing. A 3D use case for KG visualization is Redgraph by Halpin et al. [6]. In this VR application, KGs about data from the US Patent and Trademark Office are presented. The graph starts in 2D and is laid out by simulating edges as springs. After that, the user can pull out selected nodes to adjust the layout and extend it into 3D space.

3 Methodology

Our implementation imports a given RDF graph and visualizes it in the 3D space so that the learner can inspect and memorize it. The application is realized in the Unity 3D engine. To achieve high accessibility for students, it targets mobile devices. An example of the resulting application can be seen in Fig. 1c. The cross-device MR functionalities are integrated using Unity's AR Foundation package. Moreover, we utilized the Toolkit for Unity which provides features such as the 3D model import, file caching and manages the large number of node objects in the 3D scene. To read the RDF graph formats, the library dotnetRDF was utilized. This way, any RDF knowledge graph can be imported into the system to visualize it in MR. The knowledge graphs could be hand-curated similar to mind-maps or are commonly machine-generated. A possible tool which we used to gain the KG data in learning scenarios is the T-MITOCAR tool which is able to convert texts into knowledge maps in the RDF format [9]. Our resulting MR learning application is open-source and available on GitHub[1].

During the import process of the RDF graph, the system iterates over the triples that make up the graph. The subjects and objects of the triples are visualized as nodes that are connected by the predicates of the triples. The algorithm extracts suitable labels for the nodes and the edges between them based on the qualified name but the user has also the option to manually set a label with own text. Once the graph is constructed in the 3D scene, a continuous force-based layout algorithm is applied based on the algorithm by Fruchterman and Reingold [5]. Forces pull the connected nodes towards each other while repelling other nodes. A central force ensures that the graph stays centered compactly around a midpoint. This way, tightly connected sections of the graph appear as more dense clouds in the 3D representation. This helps identify closely related concepts in the graph as their semantic closeness will also be translated to the spatial neighborhood. Users can pull on individual nodes to position them and the rest of the graph will automatically adjust its layout around this external influence based on the forces. The graph's midpoint, shown as a colored sphere, can be repositioned in the room to move the entire graph. To achieve this, a special placement mode was created. It utilizes the scans of AR Foundation that identify plane surfaces in the real environment. If the user taps on such a surface, the graph's midpoint is placed 1.2 m above the point on the surface to allow for a free 3D distribution of the graph. By tapping onto a scanned flat surface, the user

[1] https://github.com/rwth-acis/MR-KGV.



(a) Sphere representation. (b) Image representation. (c) App on a smartphone.

Fig. 1. Two KG representations and a student using the app on a smartphone.

can select where the midpoint should be placed. To account for the 3D shape of the graph, the midpoint is then positioned 1.2 m above the selected surface. These AR visualizations of the graph have a series of advantages over pure screen-based 3D visualizations. First, the user has more intuitive movement controls and can, e.g., walk around the graph and explore it by walking into the network. Since the real world stays visible, it can act as an additional reference frame. The users are in familiar surroundings and complex movements like rotations can be understood better since the visual flow of the real objects in the background is observable. This can also support the user's spatial understanding to better grasp the general shape and characteristics of the shown graph. In contrast to this, virtual 3D graphs on a screen are usually shown in a digital void where users can quickly lose orientation if the nodes are too uniform or if the virtual camera looks at an empty part of the space.

Different node representations were added to the visualization to investigate their effects. In the sphere representation which is depicted in Fig. 1a, all nodes of the graph are shown as uniformly colored spheres. On the spheres, the human-readable name of the KG node is displayed. In the learning context, this usually corresponds to a concept, actor or object from the particular field of study. As the second mode, nodes can be switched to the image representation as illustrated in Fig. 1b. In this kind of visualization, the spheres are replaced with 2D images. The images are automatically fetched from an online image service based on the node's content. For instance, in a graph about neuroscience, the user can see, e.g., the node for a neuron at a glance since it is displayed with a typical depiction of a neuron. If the user has a specific image in mind for the topic of one node, the image of the node can be overwritten. For this, the user can tap onto a node which opens a settings menu. The menu allows the user to specify the Uniform Resource Locator (URL) of the image which is then subsequently downloaded and shown on the node. The third node representation works similarly to the image representation but instead of 2D images, 3D models are fetched and shown for the nodes. Both the image and 3D model representations give nodes a visual identity so that they can quickly be recognized again in the graph based on their visual appearance. The representation can also convey visual information, e.g., if the word on the label is written in a foreign language or when it is an unknown specialist term. The resulting MR experience can then be viewed on a smartphone as shown in Fig. 1c. With the spatial scan of the application, the device is able to track the user's movements through the space. So, users can reveal and explore the graph by moving around and filming their real surroundings into which the graph visualization is inserted.

4 Evaluation

We conducted a pre-study and a user study to assess the effectiveness and usefulness of the MR learning visualizations based on the KGs. It comprised a pre-questionnaire and a user evaluation, with the overall goal of collecting data to answer our research question and identify potential areas for system improvement. The questionnaire was categorized into three sections:

General Information. This section inquired about participants' gender, age, course of studies, and prior experience with MR applications. This data was essential to contextualize the results and account for any biases, such as over-representation from a specific academic course or age group.

Node Representations. Participants were introduced to the concept of a KG and then presented with three different KGs visualized using distinct representations: sphere, image, and 3D model.

Additional Feedback. This section comprised a single query seeking further feedback, particularly concerning features they would anticipate or desire in a KG visualization application.

Before the implementation, 15 participants took part in the pre-study. Here, the preferences of learners regarding different node representations in KGs were gathered based on static renders of the potential visualizations. The purpose of this study was to gain insights that could guide the implementation, e.g., about the perceived importance of each visualization option. With the final application, we held a user evaluation with 10 participants, consisting of three women and seven men. Their ages range from 24 to 58, and the majority are computer science students. Most participants have little to no prior experience with MR.

First, they received a short introduction about KGs. Participants began the session by opening the application on their device or the device provided by us and starting with the first graph and representation. We have offered two KGs that focus on neuroscience and climate change. The climate change KG demonstrated the sphere representation, whereas the neuroscience KG was displayed with the image representation. After completing the memorization and understanding task with both KGs, participants answered a LimeSurvey questionnaire. The questionnaire consisted of questions about usability and technology acceptance using System Usability Scale (SUS) [1] and Technology Acceptance Model (TAM) [3]. It also asks participants to compare and rate each node representation on a 5-point Likert scale based on the perceived suitability for learning.

In the pre-evaluation, participants ranked the representations according to their preference. The questionnaire also enquired about the reasons behind the selected order. The results showed a tendency towards the image representation due to its ability to establish quicker to understanding and clearer associations. The sphere representation also got good feedback but there were some concerns that the spheres looked too uniform. In the main user study, the usability of the application was rated with an average SUS score of 87.75, indicating 'Excellent' usability. The detailed results of the SUS and TAM statements are provided in Table 1. In the TAM, positive results were achieved which indicate that participants have a positive attitude towards the new technology. Especially the average values of the perceived ease of use, the self-efficacy and system accessibility reached scores of 4.5 or above. This is in line with the good usability result and also shows that the application has a good accessibility on the smartphones. The lowest value is the behavioral intention. Possible reasons for this include the necessity for a better integration into existing learning workflows or conveying more use cases that are closer to the participants' field of study.

In the memorization test, we compared the performance with the sphere representation against the image representation. The group had to recall connections between nodes and determine the overall topic of the KG. We awarded a total of 130 points for the answers of the entire group. On the sphere representations, they achieved 102 points and with the images, they scored 109. When asked to compare the two representations, the participants also stated a preference for the image representation. There is a trend that the content of nodes and the graph was easier to identify with images. This can also be seen as a possible reason for the improved memorability as the learners can associate the content with the shown images. The spheres were also perceived as more demanding because of their uniformity. Due to the relatively low sample size, this preference can only be seen as a trend and will need to be verified in future large-scale evaluations. In the qualitative feedback, one additional insight was that participants liked using the images as annotation tools, e.g., to highlight important nodes. Moreover, they used the combination of images and physical locations to build memory anchors with also serve as an enhancement for the memory performance. One user also noted that the application is well-suited for vocabulary studying as the words on the labels can be matched with the visual

Table 1. SUS & TAM results: Summary of means and standard deviations. ($n = 10$)

#	Question	$\bar{x}$	(σ_x)
SUS_1	I think that I would like to use this system frequently	3.8	(±1.23)
SUS_2	I found the system unnecessarily complex	1.2	(±0.42)
SUS_3	I thought the system was easy to use	4.7	(±0.48)
SUS_4	I think that I would need the support of a technical person to be able to use this system	1.7	(±0.95)
SUS_5	I found the various functions in this system were well integrated	4.5	(±0.71)
SUS_6	I thought there was too much inconsistency in this system	1.3	(±0.48)
SUS_7	I would imagine that most people would learn to use this system very quickly	4.8	(±0.42)
SUS_8	I found the system very cumbersome to use	1.3	(±0.48)
SUS_9	I felt very confident using the system	4.4	(±0.52)
SUS_{10}	I needed to learn a lot of things before I could get going with this system.	1.6	(±0.97)
TAM_{PE}	It is easy to become skillful at using this application	4.5	(±0.53)
TAM_{PU}	This application would improve my learning performance	4.2	(±1.03)
TAM_{AT}	I am positive towards using this application for visualization purposes	4.4	(±0.84)
TAM_{BI}	I intend to use this application in the future	3.9	(±1.10)
TAM_{SE}	I have the necessary skills for using this application	4.5	(±0.85)
TAM_{SA}	I have no difficulty accessing and using this application	4.6	(±0.52)

images on the node. One technical request regarded a zoom function so that the nodes can be enlarged, e.g., to inspect the images.

5 Conclusion and Future Work

In this paper, we presented a study about learning with 3D visualizations of KGs in MR. We focused on the impact of different node representations like sphere representations, images and 3D models on memorization. The results show that learners prefer to work with nodes that are visualized as images which display the content of the node. This result was both evident in the subjective comparison by the participants and in the objective memorization performance as the images allow users to recognize nodes more easily and to grasp the content. The outcome shows that the visualizations of KGs in MR bear potential for computer-supported learning activities that both give a high-level overview of the topic to students and help with detailed learning activities. Due to this trend that the learning results can be improved by working with KG visualizations, in the future work, we will focus on examining the possible learning activities and use cases in detail. For instance, the KG-based approach is suitable for topics which can be visualized well, e.g., anatomical organs in medicine, protein structures in biology or chemical compounds in chemistry. It can be examined whether conveying more abstract topics, e.g., in mathematics also benefit from this method. On the technical level, the user feedback, such as the zoom functionality and an improved UI will be added. The future research involves conducting a large-scale user study to gain further insights and to investigate factors such as the impact of display technology, e.g., by comparing head-mounted displays with the smartphone solution and 2D desktops. Moreover, we will explore the potential of the KGs in TEL to allow for computer-based processing of the data, e.g., by adding a mixed reality agent which can support the learning activity.

References

1. Brooke, J.: SUS: a quick and dirty usability scale. In: Jordan, P.W., Thomas, B., Weerdmeester, B.A., McClelland, I.L. (eds.) Usability Evaluation in Industry, pp. 189–194. Taylor & Francis (1996). https://www.taylorfrancis.com/chapters/edit/10.1201/9781498710411-35/sus-quick-dirty-usability-scale-john-brooke
2. Buschel, W., Vogt, S., Dachselt, R.: Augmented reality graph visualizations. IEEE Comput. Graph. Appl. **39**(3), 29–40 (2019). https://doi.org/10.1109/MCG.2019.2897927
3. Davis, F.: A Technology Acceptance Model for Empirically Testing New End-User Information Systems. Ph. D., Massachusetts Institute of Technology, Massachusetts (1985). http://hdl.handle.net/1721.1/15192
4. Drogemuller, A., Cunningham, A., Walsh, J.A., Ross, W., Thomas, B.H.: VRige: Exploring Social Network Interactions in Immersive Virtual Environments. https://doi.org/10.13140/RG.2.2.34201.67689
5. Fruchterman, T.M.J., Reingold, E.M.: Graph drawing by force-directed placement. Softw. Pract. Exp. **21**(11), 1129–1164 (1991). https://doi.org/10.1002/spe.4380211102
6. Halpin, H., Zielinski, D.J., Brady, R., Kelly, G.: Exploring semantic social networks using virtual reality. In: Sheth, A., et al. (eds.) ISWC 2008. LNCS, vol. 5318, pp. 599–614. Springer, Heidelberg (2008). https://doi.org/10.1007/978-3-540-88564-1_38
7. Herman, I., Melancon, G., Marshall, M.S.: Graph visualization and navigation in information visualization: a survey. IEEE Trans. Visual Comput. Graph. **6**(1), 24–43 (2000). https://doi.org/10.1109/2945.841119
8. Kortemeyer, G.: Virtual-reality graph visualization based on Fruchterman-Reingold using unity and SteamVR. Inf. Vis. **21**(2), 143–152 (2022). https://doi.org/10.1177/14738716211060306
9. Pirnay-Dummer, P., Ifenthaler, D.: Reading guided by automated graphical representations: how model-based text visualizations facilitate learning in reading comprehension tasks. Instr. Sci. **39**(6), 901–919 (2011). https://doi.org/10.1007/s11251-010-9153-2
10. Potnis, A.V., Durbha, S.S., Shinde, R.C., Talreja, P.V.: Towards visual exploration of semantically enriched remote sensing scene knowledge graphs (RSS-KGs). In: 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, pp. 5783–5786. IEEE (2021). https://doi.org/10.1109/IGARSS47720.2021.9554836
11. Schreiber, G., Raimond, Y., Manola, F., Miller, E., McBride, B.: RDF 1.1 Resource Description Framework: Primer (2014)

Tracking the Adaptive Learning Process with Topics Ontology

Martin Homola, Zuzana Kubincová(✉), Rastislav Urbánek, and Ján Kl'uka

Comenius University in Bratislava, Mlynská dolina, 84248 Bratislava, Slovakia
{homola,kubincova,jan.kluka}@fmph.uniba.sk, urbanek18@uniba.sk

Abstract. We focus on an ontological model of the learner's context in the organized education process at a university level. Extending an existing temporal context model to cover topics of study, their mutual hierarchy, and their relations with other entities in the course we yield a complex model of the learner's context. Such a model enables tracking of the student's learning goals within a course and providing personalized recommendations w.r.t. the current level of students' topic understanding as well w.r.t. the current point in time within the course. We also outline how the unified knowledge-graph representation allows for a more flexible design of visualizations and user interfaces.

Keywords: Ontology · Personalization · Learner's context

1 Introduction

One of the widely acknowledged [9] desired features of virtual learning environments (VLE) is their *adaptability*, or personalization, towards individual users' needs. The traditional view is often rooted in intelligent tutoring systems as described by Wollf et al. [16], building on three models: the *expert model* – or the learning domain model; the *student model* – capturing the learner's understanding of the learning domain and individual learning needs, and the *instructional model* – covering the organization and execution of the learning activity. These models are sometimes extended with other layers and components [4], but mostly the key functions are based on modelling of the user's state of understanding of the expert model [5], and possibly some individual user's needs (e.g. learning styles), and deriving recommendations to engage with selected learning materials or other learning objects [2]. This paradigm is thus predominantly user-centred and its key component is often referred to as user modelling [5].

While user modelling already paves ways for useful personalization features, as pointed out by Brusilovsky and Milán and by Aroyo et al., it is essential to position the learning process with respect to the "context of user's work" [5] or the "circumstances in which it occurs" [4]. As we prefer to say, the state of the user model needs to be framed within a broader *learner's context* consisting of different relevant aspects, including: user's understanding of the expert model; the

H. Xie et al. (Eds.): ICWL 2023, LNCS 14409, pp. 155–165, 2023.
https://doi.org/10.1007/978-981-99-8385-8_13

logical organization of the content in the expert model (e.g. preceding and consecutive topics); user's preferred learning styles; usage platform and paradigm (e.g. desktop vs. mobile); physical location; personal and social context; and, within organized higher education, also the progress of a course and a study program.

In our university setting, we support in-person courses with a VLE, where students find learning materials and information, they can submit assignments, and track their results. The system enables: course and user management, management of learning materials, assignments specification and submission, peer review, code review, and teamwork peer feedback, development and crowdsourcing of online quizzes including their administration, evaluation of students' work and tracking of their progress.

Previously [7] we have enriched this VLE by an ontological model explicitly representing the temporal aspects of the learner's context, capturing events framing a university course with the aim to tailor the user experience for the students. The system organizes the learning path of a student on the timeline of the course and it is able, e.g., to recommend useful study materials relevant to an upcoming test. It features a fully semantic data representation and the ontological model was published and it can be reused by other systems to support analogous features. On the other hand, our system, so far, did not feature the expert model and was not able to provide more fine-grained personalized recommendations akin to more traditional adaptive e-learning systems.

In this report, we describe how the ontological model is extended to cover the course topics space and the student's understanding of these topics. This allows us to: visualize the topics space of a course and indicate an individual student's coverage of the learning content; but also to visualize a course's topical coverage of a wider topics space in the study program. The topics space is integrated with other entities in the ontology such as assignments, quizzes and learning materials, which allows to generate more fine-grained recommendations for the students that are respective to the current context of the course and of their learning progress as well. The ontological representation captures contextual properties such as the subtopic–supertopic relation and the topic prerequisite relation which allows to further improve the orientation in the topics space but also the quality of the resulting recommendations. The fully semantic representation based on knowledge graphs has further benefits for e-learning system design – it allows us to design visualizations and user interfaces in a flexible and unified way.

2 Related Works

A number of recent studies have focused on ontology-based recommendations in educational environments [9,11,15]. While conventional recommenders are usually based on ratings, ontology-based recommender systems enhance the recommendation process with ontological domain knowledge about the learner and learning materials [13]. They are becoming popular in the field of education specifically due to their ability to personalize learner profiles based on characteristics such as their background, learning style, preferences, etc. [15].

According to Roussey et al. [12] ontologies can be categorized based on their scope into the following categories: domain ontology, application ontology, reference ontology, general ontology, and top-level ontology. Several surveys of ontology recommender systems in educational settings show that the huge majority of such systems make use of domain ontology and only a few use other ontologies, e.g. a reference ontology, a generic ontology, or a task ontology, also mostly in combination with a domain ontology [11,15].

Another view of ontology classification [3] is based on how ontologies are used in e-learning: curriculum modeling and management, describing learning domains (subject domain ontology and learning task ontology), describing learner data, and describing e-learning services. In most knowledge-based recommender systems in education, ontology is used to represent knowledge about the learner, learning objects as well as domain knowledge [8,13] in order to establish a relationship between learners and their preferences about learning resources.

Ontological educational recommender systems usually generate recommendations for learners, most commonly learning objects [11] or learning objectives [6], but also learning paths, feedback, or learning devices. However, there are also recommender systems that provide recommendations for teachers, e.g. the best teaching strategies or pedagogical scenarios to use in the context of a particular class, new resources to augment the course curriculum, etc. [6].

Several authors report the use of multiple ontologies in their systems for e-learning, most often maintained separately. Huang et al. [8] use three independent ontologies (course ontology, learner ontology, and learning object ontology) in their personalized system for recommending learning paths, learning contents, and learning experiences. The recommendations are generated using the similarity matching between the respective ontologies.

Zhuhadar and Nasraoui [17] proposed a hybrid recommender system using two domain ontologies – one for representing learning materials and the other for representing learners. They combine the content-based and the rule-based models to provide the learner with recommendations for learning concepts.

A framework for a smart e-learning ecosystem designed by Ouf et al. [10] employs four separate ontologies and a semantic reasoner to provide the learner with a personalized learning package consisting of a learner model and all components of the learning process: suitable learning objects, favorite learning activities, and the right teaching methods based on their individual preferences and needs.

Personalization is proving to be an important component in learning environments for MOOCs as well. For example, Agarwal et al. [1] propose a hybrid recommender system for MOOCs with two domain ontologies that exploits clusters-based collaborative filtering and rule-based recommendation using SWRL. The two ontologies used (learner ontology and course ontology) are designed individually, but the course ontology has been modeled with respect to the structure of the learner ontology. The system recommends individual course elements, learning paths, and general tips and suggestions about learning.

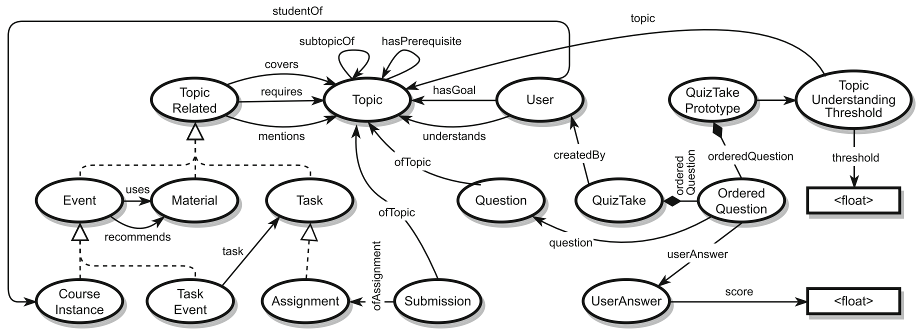


Fig. 1. Overview of the ontological model.

In contrast to the previously mentioned systems, Shishehchi et al. [13] propose a personalized knowledge-based e-learning recommendation system using a common ontology for learners and learning materials. The ontology class representing the learning material consists of two subclasses: learning topic and learning activities (exercises and quizzes). The learner class has subclasses representing the learner's learning style, profile, history, and background knowledge. Learners are recommended learning materials, activities, examples, etc., based on the learner's request and their previous knowledge.

The PLSP recommender system also proposed by Shishehchi et al. [14] provides the learner with a personalized learning sequence in the area of programming, depending on the learner's level of proficiency. The knowledge module of the system is based on a single domain ontology representing learners, learning activities, and learning materials. The recommender module built on semantic rules and ontology representation presents the learner with recommendations for learning material and learning material sequences based on the learner's request.

3 Ontological Model

Recently we have enriched our system with a semantic contextual layer that enables to organize and track the educational process within courses w.r.t. to the passing time of the course run [7]. This brings in a certain level of contextual adaptivity, e.g. recommending learning materials that are relevant at the given point in time. To further improve the adaptive capabilities we have now enriched the learner's context by integrating the representation of the topics space.

Our VLE builds purely on an integrated, fully semantic data representation, captured by a knowledge graph in the RDF format, and stored in the Virtuoso triple store. This model covers all functions and all data managed by the system. The model is rather complex; a simplified representation of the relevant part is depicted in Fig. 1. We now briefly describe the most relevant parts or the model.

Courses and events. The model stores information about courses, comprising multiple course-related events (lectures, lab sessions, tests, homework assignments, etc.). From the events within one course, the system generates a timeline which aids students' orientation in tracking the course. This part of the model and the system was covered in detail in our previous work [7].

Topics space. The core class of the current work is the Topic class. Its instances form a hierarchical topical space which stands for the *expert model*. Each topic represents a certain part of the knowledge the students may learn in one of the courses. The topics may be inter-related by two properties: subtopicOf enables to divide a topic into multiple subtopics, when it is reasonable to assume that the parent topic is completely covered by its subtopics, i.e. that mastering all subtopics implies mastering the parent topic; requires enables to indicate that this property's target topic is a prerequisite for the source topic, i.e. mastering the former is recommended before attempting the latter.

User's knowledge. The User class encapsulates all user data, most of which is not of our interest here. What is relevant is that it also enables tracking the topics which are the user's learning goals via the hasGoal property, and those which the user has achieved via the understands property.

Evaluation outputs. There are different possible ways how to actually track the user's progress towards their learning goals. The model features data about user actions from which this can be partly deduced. Once the instructors grade the user's submission to an assignment, the achievement of the goal may be deduced if the assignment covers the given topic (Assignment, Submission together with the ofTopic property). Similarly, when the user takes a quiz in the system, achievement of the goal may be deduced if the automatic evaluation of the questions on the goal's topic passes a threshold set by the instructors (QuizTake, related classes, the threshold data property).

Study materials. The Material class represents different internal or even external documents serving as learning materials. The model allows to express that a material (and also selected other entities) exhaustively covers a certain topic, non-exhaustively mentions one, or requires one (mastering it is necessary before studying the material). This enables producing recommendations for the users to study certain learning materials to achieve the learning goals.

Reflecting to the aforementioned traditional model of adaptive learning systems, the study materials and the topics would fall under the expert model, the user's data including the topic understanding would fall under the user model, and the course and its partition into the events, including assignments and quizzes would fall under the instructional model. However, we prefer to view the timeline of events and the topical annotations as a representation of the learner's context respective to a given point in the course run together with the individual point on the learner's path.

4 Model Population and System Functions

In order to make use of all the rich semantic information specified in the model the key questions concern (a) how to populate the model with actual data, and (b) how to make use of the data in the system to the benefit of the users.

4.1 The Topics Space

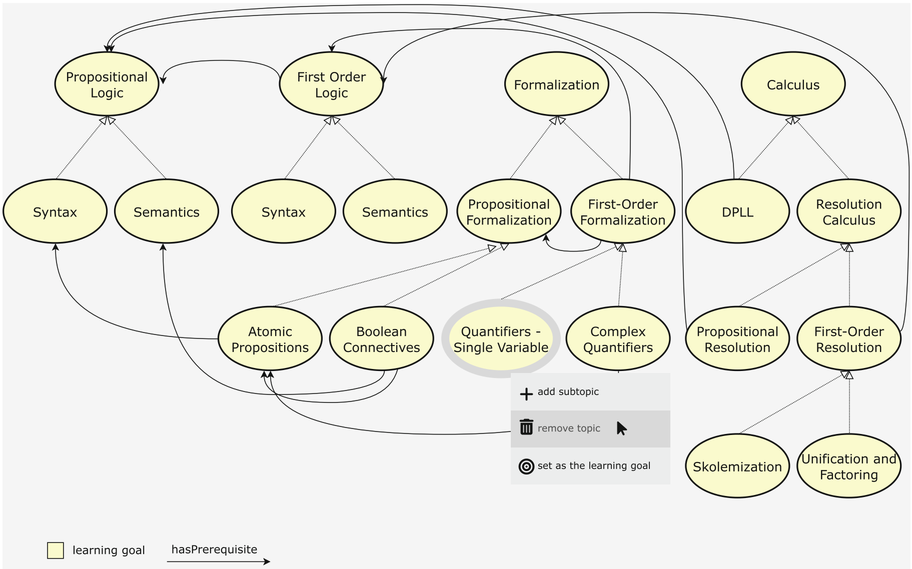


Fig. 2. Instructor's view of the topics user interface.

Creating a meaningful and sufficiently granular hierarchy of topics is critical to the overall success of the system. For the purpose of the deployment of the system within a single university or a study program, we assume:

– The hierarchy of topics will be global and shared by all courses. This will enable to visualize how a certain course fits within a global hierarchy, which courses cover the required topics a student should master before enrolling in another course, etc.
– Topics will be created and updated collaboratively by a group of instructors. We assume manual editing (but we do not rule out automatic import from existing ontologies in the future).
– An instructor will be able to set learning goals for each individual course by marking certain topics in the global hierarchy as goal topics of that course. As we assumed that a supertopic is completely covered by its subtopics, the goal topic's subtopics automatically become goal topics of the course.

The two latter function will be enabled by a unified instructor's interface sketched in Fig. 2.

4.2 User Model

For the model of student knowledge, there are potentially three ways of building and updating the data:

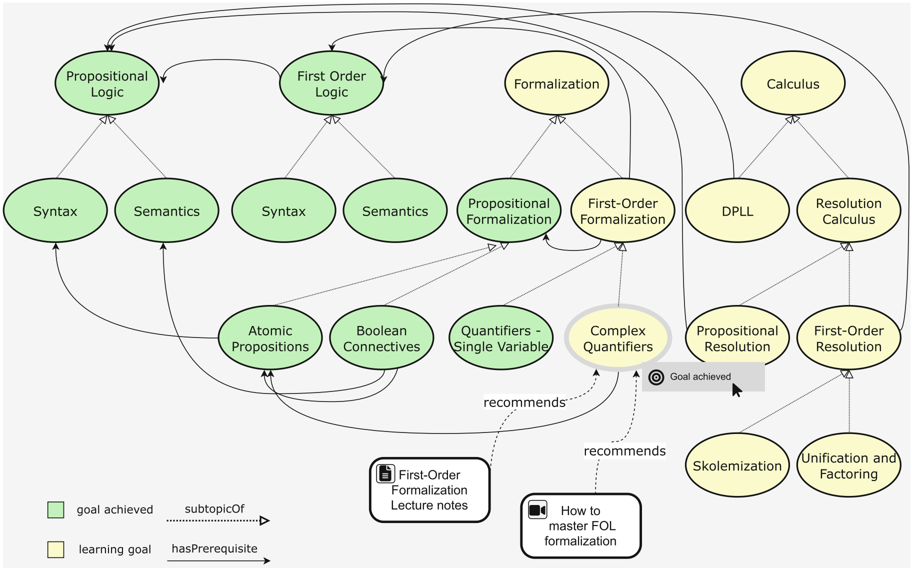


Fig. 3. Student's view of the course topics space.

1. *Self-directed:* The students manually mark the mastery of a given topic in the course's topic hierarchy view. This option is visible in the student's overview of the course topics in Fig. 3.
2. *Informal:* Topic mastery is updated automatically based on informal formative tasks (exemplified in our model by quizzes) prepared by the instructors. Each quiz may cover one or multiple topics and for each of them it may specify a *topic understanding threshold*. If a student's overall score from the questions on a given topic surpasses the threshold, the topic is marked as understood by the student.
3. *Formal:* The topic mastery is updated automatically, in a similar fashion as in the informal mode, however, the thresholds are defined over formal evaluation points awarded to the students by instructors for the submitted formal assignments.

The three modes of building the student's knowledge model will be configurable by the instructors depending on which one they want to use in their course. We assume that modes can also be combined. Creating such a model will enable the system to offer the following functions:

– Tracking the student's progress (for both the students and the instructors) by viewing the topics marked as understood by the given student in the course topic hierarchy view. The student's version of the interface is depicted in Fig. 3. Thanks to the assumption that a supertopic is completely covered by its subtopics, the goal of the supertopic is automatically achieved once the student understand all of its subtopics. This assumption is useful to avoid

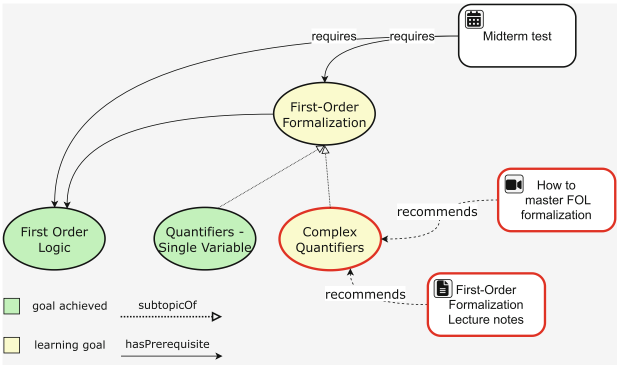


Fig. 4. Recommendation view.

confusing situations. Students can also activate any topic, see its related study materials, and mark it as understood (in the self-directed mode).

- The state of the student's topic mastery within the course, together with the overall context of the course may be used for adaptive purposes such as to produce individualized recommendations of study materials to follow at a given moment. An example is shown in Fig. 4 where shortly before the midterm test the student is given an overview of the topics required by the test with a recommendation of study materials covering the topics the student has not mastered yet.
- The student's individual knowledge model may also be visualized w.r.t. the global topics hierarchy. This would also enable to produce recommendations at the level of courses, such as which course to take next to fill in a student's gap in a certain area of the topics.

We have discussed how the proposed ontological model needs to be maintained, but more importantly, how it allows to provide new useful functions to the benefit of the users (both students and instructors).

5 Universal User Interface

Representing the data in a knowledge graph based on a well-defined ontological model also enables to define more flexible and more universal user interfaces, as explained in this section.

The interfaces in Figs. 2, 3 and 4 all visualize individual entities that exist as objects in our knowledge graph. Therefore for building each view in the user interface, it is essentially required:

1. to obtain the entities of interest together with the links that we want to include from the knowledge graph;
2. to define an appropriate visualization style (shape, colour, etc.) for each entity based on its type;

3. to define possible actions for each entity based on the view, entity type, and user level; and finally
4. to visualize them using a suitable graph visualization tool.

For example, the topics related to a given course, together with their interlinks respective to the subtopicOf and hasPrerequisite relations may be obtained using the following query:[1]

```
CONSTRUCT {
  :t1 a :Topic. :t2 a :Topic. :t3 a :Topic. :t4 a :Topic.
  ?t1 :subtopicOf ?t2.
  ?t3 :hasPrerequisite ?t4
} WHERE {
  ?t1 :subtopicOf ?t2.
  ?t3 :hasPrerequisite ?t4.
  :t1 a :Topic. :courseXY :covers ?t1.
  :t2 a :Topic. :courseXY :covers ?t2.
  :t3 a :Topic. :courseXY :covers ?t3.
  :t4 a :Topic. :courseXY :covers ?t4
}
```

In order to generate the student's view in which topics that the student already achieved as their learning goal are distinguished (i.e. marked in green in Fig. 3), it suffices to obtain these topics and mark them by a distinctive class. To this effect, the following query is run and the resulting graph is merged with the one obtained by the previous query by the UNION operation:

```
CONSTRUCT {
  :t a :CoveredTopic
} WHERE {
  :t a :Topic.
  :courseXY :covers :t.
  :studentYZ :understands :t
}
```

Here the class :CoveredTopic is not part of our underlying model but is merely used for visualization purposes.

6 Conclusions

We have described an ontological model that we adopted in our VLE in the scope of organized education at our university. Specifically, we have focused on extending the model to cover the course topics space, its hierarchical organization, and

[1] For simplicity of presentation we assume all entities to pertain to the default namespace ":" and that :courseXY is the individual respective to the current course instance of interest in this view. In the following query, :studentYZ is the individual respective to the user (or the student whose progress the teacher wishes to visualize).

its integration with other relevant entities of the model. This extension allows us to greatly improve the representation of the learner's context respective to the point of time within a course run but jointly also to the student's level of understanding of the covered matter.

We have consecutively presented novel features enabled by this extension. This includes the organization of the topics hierarchy and using it to track the learning goals of a course and the progress of an individual student. But thanks to the tight semantic integration with other stored data in a unified contextual layer, it now especially allows us to provide much more sophisticated personalized recommendations for individual students. Finally, we have discussed how the unified knowledge-graph representation allows for a more effective and flexible design of user interfaces.

Our approach differs from what has been reported in the literature in several respects:

- We build one comprehensive ontology that covers all data stored and manipulated by the system (thus a complex application ontology rather than just isolated domain ontologies used for recommendations).
- We focus on organized courses of formal education, in contrast to many studies which focus on self-directed e-learning scenarios. Therefore we need to account for courses, learning sessions, evaluation activities, grading, and their organization in time.
- To achieve the above, our ontology provides a model of an integrated learner's contextual space connecting the topical and the temporal aspects.
- Tracking of the student's position in this contextual space allows us to proactively recommend relevant learning activities and materials at any given time without the learner needing to request them from the system.
- Our unified knowledge-graph representation offers other benefits, such as flexible visualization and user interface generation.

In the future, we would like to further extend the richness of the model of the learner's contextual space, e.g. by accounting for students' individual needs, such as learning styles, access modality (desktop or mobile), but also a richer notion of access circumstances (during individual study time, during the lecture or lab session, on the go, etc.), and possibly others.

Acknowledgements. The work was supported by national projects nos. VEGA-1/0621/22 and APVV-20-0353 awarded by Slovak VEGA and APVV research agencies.

References

1. Agarwal, A., Mishra, D.S., Kolekar, S.V.: Knowledge-based recommendation system using semantic web rules based on learning styles for MOOCs. Cogent Eng. **9**(1), 2022568 (2022)
2. Akbulut, Y., Cardak, C.S.: Adaptive educational hypermedia accommodating learning styles: a content analysis of publications from 2000 to 2011. Comput. Educ. **58**(2), 835–842 (2012)

3. Al-Yahya, M., George, R., Alfaries, A.: Ontologies in e-learning: review of the literature. Int. J. Softw. Eng. Appl. **9**(2), 67–84 (2015)
4. Aroyo, L., et al.: Interoperability in personalized adaptive learning. J. Educ. Technol. Soc. **9**(2), 4–18 (2006)
5. Brusilovsky, P., Millán, E.: User models for adaptive hypermedia and adaptive educational systems. In: Brusilovsky, P., Kobsa, A., Nejdl, W. (eds.) The Adaptive Web. LNCS, vol. 4321, pp. 3–53. Springer, Heidelberg (2007). https://doi.org/10.1007/978-3-540-72079-9_1
6. Fraihat, S., Shambour, Q.: A framework of semantic recommender system for e-learning. J. Softw. **10**(3), 317–330 (2015)
7. Homola, M., Kl'uka, J., Kubincová, Z., Marmanová, P., Cifra, M.: Timing the adaptive learning process with events ontology. In: Herzog, M.A., Kubincová, Z., Han, P., Temperini, M. (eds.) ICWL 2019. LNCS, vol. 11841, pp. 3–14. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-35758-0_1
8. Huang, C., Liu, L., Tang, Y., Lu, L.: Semantic web enabled personalized recommendation for learning paths and experiences. In: Zhu, M. (ed.) ICCIC 2011. CCIS, vol. 235, pp. 258–267. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-24022-5_43
9. Javed, U., Shaukat, K., Hameed, I.A., Iqbal, F., Alam, T.M., Luo, S.: A review of content-based and context-based recommendation systems. Int. J. Emerg. Technol. Learn. (iJET) **16**(3), 274–306 (2021)
10. Ouf, S., Abd Ellatif, M., Salama, S.E., Helmy, Y.: A proposed paradigm for smart learning environment based on semantic web. Comput. Hum. Behav. **72**, 796–818 (2017)
11. Rahayu, N.W., Ferdiana, R., Kusumawardani, S.S.: A systematic review of ontology use in e-learning recommender system. Comput. Educ.: Artif. Intell. **3**, 100047 (2022)
12. Roussey, C., Pinet, F., Kang, M.A., Corcho, O.: An introduction to ontologies and ontology engineering. In: Roussey, C., Pinet, F., Kang, M.A., Corcho, O. (eds.) Ontologies in Urban Development Projects. Advanced Information and Knowledge Processing, vol. 1, pp. 9–38. Springer, London (2011). https://doi.org/10.1007/978-0-85729-724-2_2
13. Shishehchi, S., Banihashem, S.Y., Zin, N.A.M., Noah, S.A.M., Malaysia, K.: Ontological approach in knowledge based recommender system to develop the quality of e-learning system. Aust. J. Basic Appl. Sci. **6**(2), 115–123 (2012)
14. Shishehchi, S., Zin, N.A.M., Seman, E.A.A.: Ontology-based recommender system for a learning sequence in programming languages. Int. J. Emerg. Technol. Learn. (Online) **16**(12), 123 (2021)
15. Tarus, J.K., Niu, Z., Mustafa, G.: Knowledge-based recommendation: a review of ontology-based recommender systems for e-learning. Artif. Intell. Rev. **50**, 21–48 (2018)
16. Wolff, A.S., Bloom, C.P., Shahidi, A.K., Rehder, R.E.: Using quasi-experimentation to gather design information for intelligent tutoring systems. In: Facilitating the Development and Use of Interactive Learning Environments, pp. 21–49. CRC Press (2020)
17. Zhuhadar, L., Nasraoui, O.: A hybrid recommender system guided by semantic user profiles for search in the e-learning domain. J. Emerg. Technol. Web Intell. **2**(4) (2010)

Computer Support for Intelligent Tutoring

Prompting Large Language Models to Power Educational Chatbots

Juan Carlos Farah[1,2(✉)], Sandy Ingram[2], Basile Spaenlehauer[1], Fanny Kim-Lan Lasne[1], and Denis Gillet[1]

[1] École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland
{juancarlos.farah,basile.spaenlehauer,fanny.lasne,denis.gillet}@epfl.ch
[2] University of Applied Sciences (HES-SO), Fribourg, Switzerland
{juancarlos.farah,sandy.ingram}@hefr.ch

Abstract. The recent rise in both popularity and performance of large language models has garnered considerable interest regarding their applicability to education. Technologies like ChatGPT, which can engage in human-like dialog, have already disrupted educational practices given their ability to answer a wide array of questions. Nevertheless, integrating these technologies into learning contexts faces both technological and pedagogical challenges, such as providing appropriate user interfaces and configuring interactions to ensure that conversations stay on topic. To better understand the potential large language models have to power educational chatbots, we propose an architecture to support educational chatbots that can be powered by these models. Using this architecture, we created a chatbot interface that was integrated into a web application aimed at teaching software engineering best practices. The application was then used to conduct a case study comprising a controlled experiment with 26 university software engineering students. Half of the students interacted with a version of the application equipped with the chatbot, while the other half completed the same lesson without the chatbot. While the results of our quantitative analysis did not identify significant differences between conditions, qualitative insights suggest that learners appreciated the chatbot. These results could serve as a starting point to optimize strategies for integrating large language models into pedagogical scenarios.

Keywords: Educational Chatbots · Prompting · GPT-3 · Large Language Models · Software Engineering Education · Digital Education

1 Introduction

As large language models (LLMs) become more accessible through publicly available application programming interfaces (APIs), the potential these models have to support a wide variety of pedagogical scenarios is becoming evident. One obvious application of these models is to power educational chatbots. Indeed, researchers have advocated for the use of LLMs to support more natural conversation and overcome the limitations of rule-based systems or systems based

H. Xie et al. (Eds.): ICWL 2023, LNCS 14409, pp. 169–188, 2023.
https://doi.org/10.1007/978-981-99-8385-8_14

on limited training data [15]. However, several challenges limit the integration of LLM-powered chatbots into educational contexts. These challenges include—among others—providing the appropriate user interfaces and configuring the LLMs to ensure that the generated text is aligned with the pedagogical scenario in which the educational chatbots are deployed [11].

To help address these challenges, we designed an architecture that developers in education can follow to create interfaces that educators can use to configure and deploy LLM-powered chatbots for use in digital education. At its core, this architecture considers chatbots as configuration objects that define the user interface elements that will be shown to the learner as well as the prompt that will be sent to the LLM. Following this architecture, we created a chatbot interface that was embedded in a web application designed to execute and review snippets of code. We used this application in a case study consisting of a between-subjects controlled experiment with 26 software engineering students. While 13 students completed a lesson on programming best practices supported by the chatbot, the other 13 students completed the same lesson without support from the chatbot. Our mixed-method analysis of the data from this experiment focused on how the chatbot affected five aspects of the learning experience. A sixth aspect—*conversation*—was applicable only to the treatment group that was exposed to the chatbot.

Given the increasing interest in the applications of powerful natural language processing (NLP) technologies, our study is timely and relevant to both research and practice. Our architecture can serve developers in education looking to integrate chatbots into digital education platforms, while findings from our case study can guide researchers and educators in conducting future empirical studies and incorporating educational chatbots into their practice.

2 Background and Related Work

Our research is underpinned by advances in NLP methods and, in particular, by the success of LLMs. LLMs are used for a wide array of language tasks and can also be used to power chatbots. The most notable example is OpenAI's *ChatGPT* [13]. Without further configuration, this open-domain chatbot can interact with users on diverse topics through unrestrained conversations. However, understanding how LLMs can be harnessed to create *task-oriented* chatbots for specific domains is still the focus of ongoing research.

A recent position paper by Kasneci *et al.* [11] outlined opportunities and challenges of incorporating LLMs into education. Among the challenges identified, the authors highlight (i) the possible biases that LLMs can perpetuate and amplify, (ii) the need for open educational resources (OERs) to guide educators on how to access and use these models, (iii) the need to ensure data privacy and security, and (iv) the lack of adaptability to align the models with the objectives of individual learners and educators. Addressing these challenges could open the door to more powerful applications of LLMs to education. Kasneci *et al.* also underlined the need to ensure that the interfaces used to interact with these models are aligned with the needs of different types of learners (e.g., adapting for age-related constraints and accessibility requirements).

An important factor to consider when assessing the applicability of these LLMs to education is the process through which they can be configured for specific tasks, such as generating the responses that an educational chatbot can use in its interactions. This process is supported by *prompting*, which consists in providing the model with instructions and a few examples of how it should respond to a query. Finding the most appropriate prompts is a complex task and using inefficient prompting strategies can result in worse results than using no prompt at all [14]. Identifying optimal prompts—or *prompt engineering*—is an active area of research in NLP. Recent work has focused on automating the generation of these prompts [9], understanding the biases that prompting can be susceptible to [18], and providing the appropriate infrastructure to generate prompts [2]. Nevertheless, there is little guidance for performing prompt engineering with domain-specific applications in mind, as is the case with educational chatbots.

These gaps in the research motivate the design of our architecture, which aims to help developers in education provide interfaces for educators to easily configure and deploy chatbots within digital education platforms.

3 Design

Our architecture—depicted in Fig. 1—consists of six components that interact through three main processes. These components are loosely coupled and abstractly defined in order to allow developers to adapt them as needed for the particular constraints that might be present in different digital education platforms. In this section, we detail the functionalities provided by each component and the processes they support.

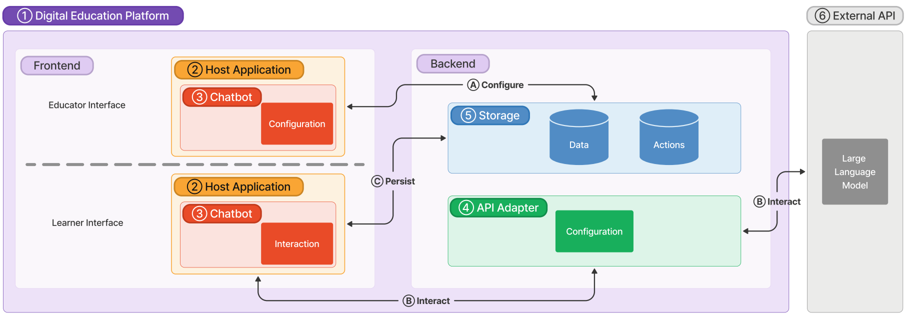


Fig. 1. Our design comprises six loosely coupled components (labeled 1–6) that interact through three main processes (labeled A–C).

3.1 Components

The architecture defines six components that interact to facilitate the integration of chatbots powered by LLMs into educational contexts: (1) digital education platform, (2) host application, (3) chatbot, (4) API adapter, (5) storage, (6) external API.

1. **Digital Education Platform:** The learning activity takes place on a digital education platform. This platform should provide a way to differentiate content that is visible to educators from content that is visible to learners. Using a dedicated view, educators should be able to create a learning activity by curating multimedia content, such as images, text, and videos. Most importantly, educators should be able to select interactive learning resources (e.g., web applications) that can host chatbots. That is, chatbots are not directly embedded in the digital education platform, but in interactive applications that can be added to the learning activity. We refer to these applications as *host applications.*
2. **Host Application:** The host application is an interactive learning resource that can be embedded in digital education platforms to support learner interactions with chatbots. A host application exposes a configuration panel for educators to activate and configure chatbots. Once activated, these chatbots are visible to learners interacting with the application. For this, the host application should provide an interface through which learners can communicate with the chatbot (e.g., a forum, chat box, or message thread). How a chatbot interacts with learners is entirely defined by its configuration and the interaction affordances provided by the host application.
3. **Chatbot:** A chatbot in our architecture is not an application by default, but a component that can be activated and configured *within* an application. This increases the portability and customizability of a chatbot, which can be developed independently from the host application(s) it can be embedded in, and by educators without technical backgrounds. Hence, the chatbot can be considered to be an OER that is defined by its configuration. This configuration consists of a name, an avatar, a scope (i.e., who can see the messages sent by individual students), a prompt (which is sent to the LLM), and a cue (i.e., a message displayed to learners to invite them to interact).
4. **API Adapter:** The API adapter is embedded in the digital education platform and serves to communicate with the API exposed by the LLM provider. This adapter can also be deployed outside of a single digital education platform, thus serving multiple platforms. Given that adapters interface with APIs external to the digital education platform, they can also provide fallback responses in case the APIs are not accessible. Nevertheless, adapters are stateless and should only serve to handle requests from host applications, thus delegating the storage of any information to the digital education platform.
5. **Storage:** Two types of data are stored by the digital education platform: (i) application data and (ii) application actions. Application data refer to the content of a learner's interaction with the chatbot, including both the

messages provided by the learner and the chatbot's responses. Application actions refer to activity traces of how a learner interacts with the conversational interface in which the chatbot is embedded, including keystrokes, clicks, and other events that can be captured by the browser. These actions can serve to deliver learning analytics and provide a more nuanced depiction of how learners interact with chatbots.

6. **External API:** Finally, our architecture requires an LLM that is accessible through a public API in order to generate the responses to the queries provided by the learners.

3.2 Process

There are three main processes that define the interactions between the aforementioned components. These processes are outlined below.

(A) **Configure:** Educators should be able to configure the chatbot through the educator interface of the host application. The chatbot's configuration is kept in the storage component and allows for the personalization of the chatbot integration for each educator's particular needs. This configuration is also seamlessly reproducible, allowing educators to quickly replicate chatbot integrations across learning activities.

(B) **Interact:** The core process defined by our architecture is how learners interact with the chatbot. When a learner interacts with the chatbot, the host application's learner interface connects to an external LLM API provider through the API adapter component hosted in the digital education platform's backend.

(C) **Persist:** As interactions take place, the host application ensures that the outcomes of these interactions are *persisted* on the digital education platform. These outcomes mainly concern the conversation between the chatbot and the learner (and any related content such as emoji reactions) but also include any actions that the learner might take using the chatbot interface. These actions can then be used to provide learning analytics.

4 Methodology

To test the applicability of our architecture in practice, we implemented a chatbot integration following our design and conducted an evaluation focused on addressing one main research question:

> *How does incorporating a large language model-based chatbot to support a lesson on software engineering best practices affect the learning experience?*

We conducted this study in January 2023. This evaluation took the form of a case study consisting of a between-subjects controlled experiment comprising one control (no chatbot) and one treatment (chatbot) condition. For both conditions, we analyzed five aspects of the learning experience: (i) *short-term learning* *gains*, (ii) *engagement*, (iii) *self-reflection on the learning experience*, (iv) *feedback regarding the lesson*, and (v) *usability*. For the treatment group, we also analyzed a sixth aspect (vi) *conversation*, which focused on the exchanges the learners had with the chatbot. The small-scale nature of our study allowed us to conduct a mixed-method analysis. In this section, we present our methodology.

4.1 Scenario

To ensure ecological validity, our evaluation took place in a formal education setting. As part of their coursework, students completed an in-class online lesson consisting of an ungraded 45-min exercise. The exercise comprised a *code review notebook* [5] covering JavaScript code style standards. Code review notebooks allow educators to scaffold pedagogical scenarios that introduce the code review process to learners through code snippets, following a template resembling computational notebooks. More specifically, we started by introducing the concept of *linting* code, which involves the use of static analysis tools to detect issues in software [10]. The lesson then covered ESLint [17], a linter for JavaScript [16], as well as the Airbnb JavaScript Style Guide [1], a popular configuration for ESLint. In this section, we outline the technological context and pedagogical scenario used in our evaluation.

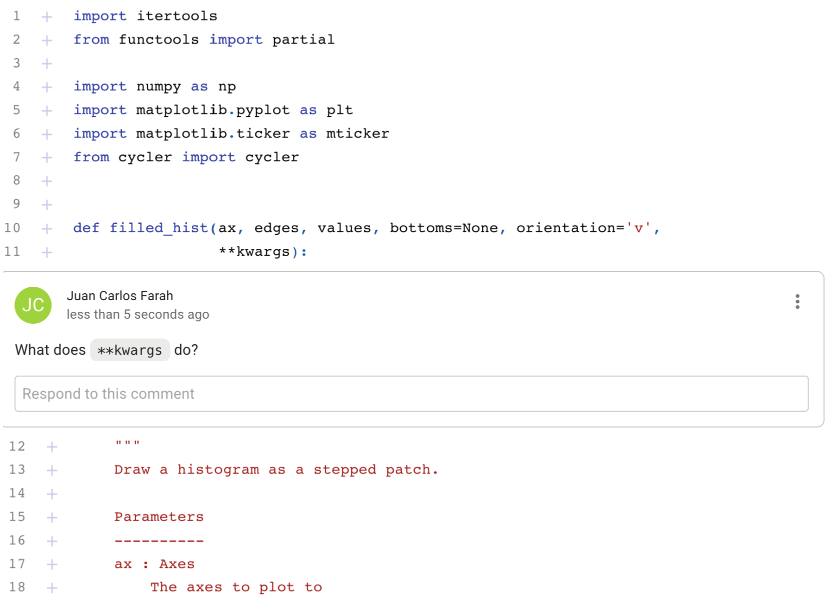


Fig. 2. The Code Capsule application can be used to write, execute, and review code. The code used in this example has been adapted from the Python Matplotlib library's documentation [7].

Technological Context. In this section, we describe how we implemented our architecture to provide the technological context for our experiment. Our chatbot was integrated into the *Graasp* digital education platform [6] through *Code Capsule* as the host application. Code Capsule is an application that allows learners to both review and execute code (see Fig. 2). This application supports the integration of chatbots when used to review code.

The chatbot that was integrated into Code Capsule was configured as follows. To remain gender-neutral and maximize consistency with Graasp, we named our chatbot *Graasp Bot* and represented it with a robot avatar. All chatbot interactions were scoped at the individual level, so students could only see their own interactions and not those of their peers. The prompts for each code snippet were only visible to the educator via Code Capsule's configuration panel and were prepended to the conversation as learners interacted with the chatbot. These prompts were defined following the pattern below[1]:

> *The following is a conversation between a chatbot and a student discussing the correction of an exercise about linting, ESLint, code styling, and best practices in JavaScript. After each response, the chatbot gives the student one or two options to continue the conversation.*
>
> ***Chatbot:***
> *`<EXPLANATION OF WHAT IS WRONG IN THE CODE SNIPPET>`. So, I would change the following line of code:*
>
> *`<INCORRECT CODE>`*
>
> *To the following:*
>
> *`<CORRECT CODE>`*
>
> *`<QUESTION ASKING THE STUDENT IF THEY UNDERSTAND THE DIFFERENCE>`*
>
> ***Student:*** OK. `<QUESTION ASKING THE CHATBOT SOMETHING RELATED TO THE ISSUE>`
>
> ***Chatbot:*** `<ANSWER>`. Do you want me to provide you with an example of *`<THE ISSUE>`*?
>
> ***Student:*** No, that's OK. Thanks! Any other issues I should be aware of?

To invite users to interact with the chatbot, we showed them a message that was embedded as a comment in the code snippet featured on Code Capsule. We refer to these messages as *cues*. Cues were dependent on each code snippet in

[1] Note that placeholders are presented between angle brackets (`<>`).

which the chatbot was embedded. All of these cues followed the same pattern (see Fig. 3). This pattern consisted of (i) an explanation of an issue present in the code snippet, (ii) a suggestion on how to fix the issue, and (iii) a question asking students whether they agreed with some facet of the proposed solution.

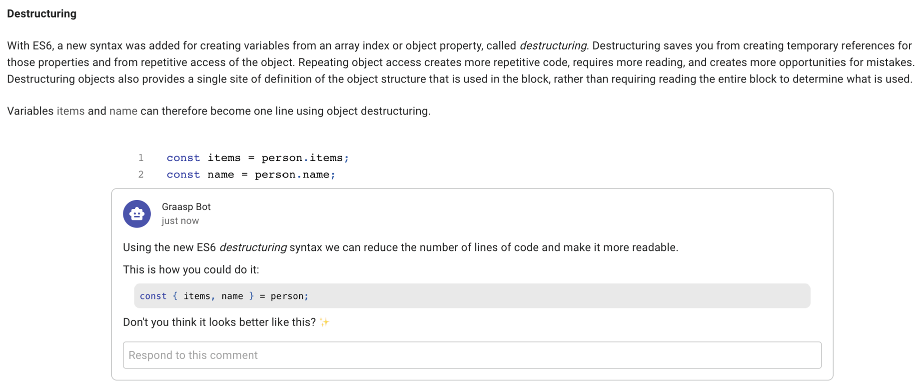


Fig. 3. The cue consisted of an explanation of the issue present in the snippet, a suggested fix, and a question asking students whether they agreed with the solution proposed.

When students interacted with our chatbot, Code Capsule interfaced with the external API exposed by OpenAI through an API adapter that was integrated into Graasp. The adapter was configured to use the GPT-3 `text-davinci-003` model, with a `temperature` parameter of 0.9 (higher values make completions of the same prompt more random), a `presence_penalty` of 0.6 (higher values penalize new tokens if they have already appeared), a `top_p` parameter of 1 (always returning the best completion), and the maximum number of tokens to be included in the chatbot response fixed at 150.

Pedagogical Scenario. The lesson used in this experiment followed the *Fixer Upper* pedagogical pattern [3]. For coding exercises, this pedagogical pattern consists of presenting students with code "that is generally sound but [contains] carefully introduced flaws [that] can both introduce a complex topic early and serve as a way to introduce error analysis and correction" [3]. In our case, we structured the lesson over 10 phases, which students were meant to navigate sequentially (see Fig. 4). Students were first introduced to the lesson (Phase 1) and then asked to complete a short exercise that served as a pre-test to gauge their knowledge of JavaScript code style standards (Phase 2). Phases 3 (*Introduction*) and 4 (*ESLint*) covered the concept of code linting and ESLint specifically, while Phases 5 (*Styling*) and 6 (*Best Practices*) presented examples of code style standards that students should follow when writing JavaScript code. In these phases, 10 code snippets were included using Code Capsule alongside a textual explanation of the issue present in the snippet. For students in the treatment condition, the code snippet also included a cue from Graasp Bot, as shown in Fig. 3. Phase 7 (*Exercise*) consisted of an exercise that served as a post-test, while Phase 8 (*Solutions*) presented the solutions to the exercise and asked the students to reflect on their performance in the exercise. For students in the treatment condition, Phase 8 also included explanations presented by Graasp Bot. Phase 9 (*Chatbots*) was also different between the control and treatment conditions. On the one hand, students in the control condition were asked to imagine how they would integrate chatbots into the exercise and to provide sample dialogs that they would envision having with the chatbot. On the other hand, students in the treatment condition were asked to report on their experience interacting with the chatbot. Finally, Phase 10 (*Conclusion*) served as a conclusion to the exercise and provided a link to a questionnaire.

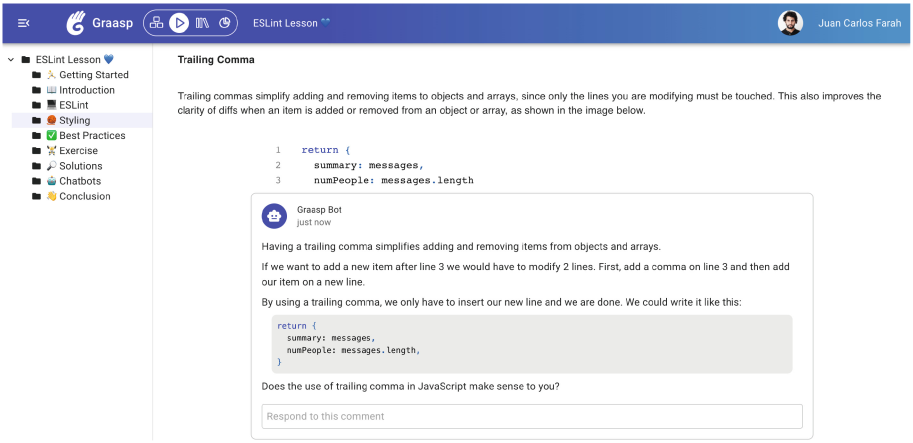


Fig. 4. The lesson used in this study consisted of a code review notebook aimed at teaching software engineering best practices and included the ten phases shown in the sidebar. In this figure, we highlight the *Styling* phase and show the explanation block—including a cue from the chatbot—recommending the use of trailing commas.

4.2 Participants

We recruited 28 third-year bachelor students taking part in a course on human-computer interaction at the School of Engineering and Architecture of Fribourg, Switzerland. A total of 26 students—25 male, 1 female—completed the study. Students were informed that this was an ungraded, optional exercise. To encourage participation, students were given extra credit for completing the activity.

4.3 Instruments

Short-term learning gains were operationalized based on the learner's performance in the pre- and post-tests by calculating the difference between both tests. These gains could range from −100% to 100%. Engagement was measured by calculating the total amount of time spent in each phase during the one-hour time frame that was allocated for the lesson. Self-reflection and feedback were respectively operationalized through the following two open-ended questions: (i) *Did you manage to find all of these [issues]? If not, which ones did you miss? Did you find any of them particularly tricky/helpful?* and (ii) *What did you think about this lesson? Any comments, suggestions, or feedback?*. A second feedback question was relevant only to the students in the treatment group: *In a few phrases, describe your experience interacting with the chatbot used in this activity. What did you like about it? What could be improved?* This second question was only used for our qualitative analysis of the feedback aspect. Usability was measured with the User Experience Questionnaire (UEQ), a standard instrument that measures usability across six dimensions [12]. Finally, for students in the treatment group, our analysis included the conversations students had with the chatbot. This qualitative analysis focused on (i) whether conversations were on topic, (ii) how long the conversations were, (iii) how natural the conversations were, and (iv) what responses were elicited by the different types of cues.

4.4 Data Analysis

We applied both descriptive and inferential statistics to our quantitative data, reporting the means ($\bar{x}$), medians ($\tilde{x}$), standard deviations (s_x), minima ($x_{\min}$), and maxima ($x_{\max}$), as well as the results of t-tests for independent samples comparing across the two conditions. To perform sentiment analyses on students' self-reflection and feedback responses, we used VADER [8], which assigns a sentiment score ranging from −1 (negative sentiment) to +1 (positive sentiment). The results of the UEQ were analyzed using its data analysis toolkit [12]. Finally, open-ended responses were analyzed using qualitative methods, following line-by-line data coding [4].

5 Results

In this section, we present our results with respect to the aspects studied.

5.1 Learning Gains

The mean learning gains were $\bar{x} = 0.429$ ($\tilde{x} = 0.369$, $s_x = 0.212$, $x_{\min} = 0.115$, $x_{\max} = 0.746$) in the control condition and $\bar{x} = 0.430$ ($\tilde{x} = 0.492$, $s_x = 0.243$, $x_{\min} = -0.146$, $x_{\max} = 0.692$) in the treatment condition. These results—illustrated in Fig. 5—show that both conditions led to, on average, positive learning gains for students, with all students except one achieving positive learning gains. While the median learning gain was higher in the treatment condition, a t-test did not show a significant difference between conditions.

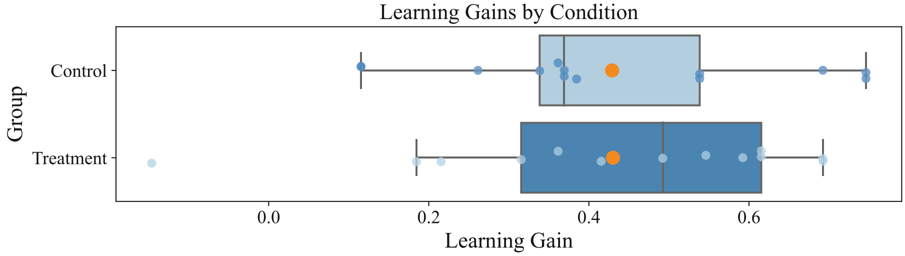


Fig. 5. Learning Gains

5.2 Engagement

On average, learners in the treatment condition spent a total of $\bar{x} = 41.7$ min ($\tilde{x} = 37.8$, $s_x = 9.31$, $x_{\text{min}} = 29.4$, $x_{\text{max}} = 59.5$) to complete the lesson, while learners in the control condition did so in $\bar{x} = 40.1$ min ($\tilde{x} = 36.5$, $s_x = 10.4$, $x_{\text{min}} = 22.2$, $x_{\text{max}} = 55.2$). There were no significant differences in the time spent by students either overall (see Fig. 6) or across the 10 phases that constituted our lesson (see Fig. 7).

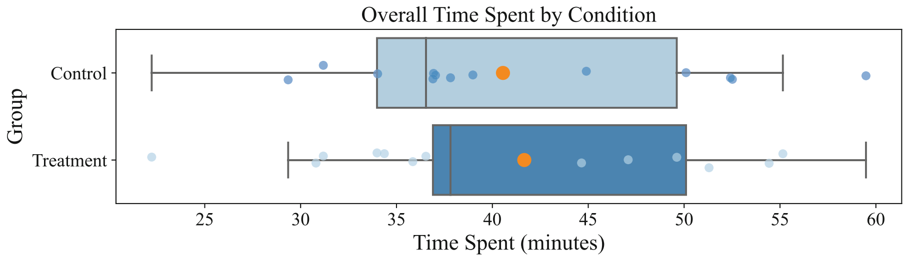


Fig. 6. Engagement

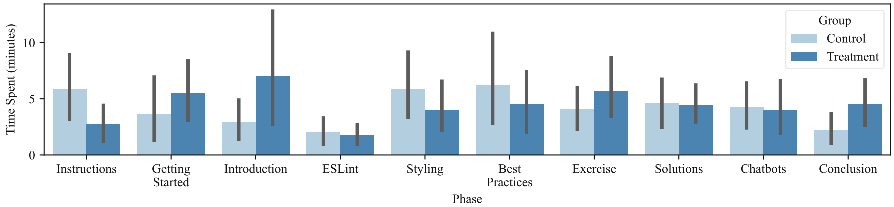


Fig. 7. Time Spent per Phase

5.3 Self-reflection

Of the 26 students that participated in the study, 22 students (10 control, 12 treatment) provided responses to the self-reflection question. The sentiment analysis performed on student responses to the self-reflection question did not produce any significant differences between conditions. Nevertheless—as shown in Fig. 8—the distribution of scores in the treatment condition had a more positive tendency than the scores in the control condition. Specifically, the responses of the students in the control group resulted in a mean sentiment score of $\bar{x} = -0.128$ ($\tilde{x} = -0.0766$, $s_x = 0.454$, $x_{\min} = -0.743$, $x_{\max} = 0.757$), while those in the treatment condition resulted in a mean score of $\bar{x} = 0.0297$ ($\tilde{x} = 0.0386$, $s_x = 0.423$, $x_{\min} = -0.595$, $x_{\max} = 0.649$).

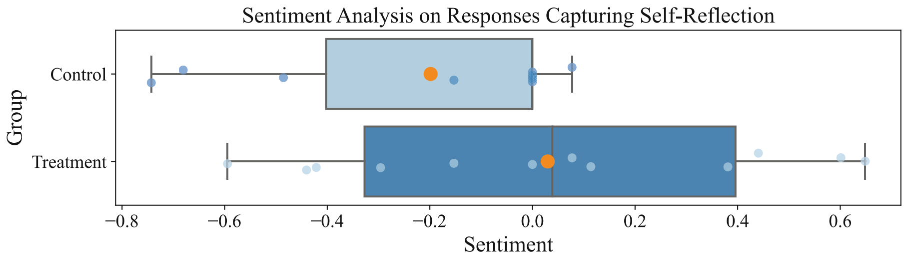


Fig. 8. Self-Reflection

Our qualitative analysis showed that responses were consistent between both groups, with students providing short answers in which they quickly described what they missed. All students, except five (two control, three treatment), specifically listed at least one issue they missed. A typical answer reads as follows: *"I forgot a `let` that had to be `const` , a comma after the last element of an object and a semicolon at the end of a line"*.

Six students (four control, two treatment) provided more detail regarding the issues they missed. In this case, a typical answer was: *"I forgot the first `let giftList` to `const giftList` despite seeing it for `total` at the end. I fell into the trap thinking it was redeclared in the `getGiftsTotal(person)` function because it had the same name"*.

5.4 Feedback

Of the 26 students that participated in the study, 22 students (10 control, 12 treatment) provided responses to the feedback question. The sentiment analysis performed on these responses did not yield any significant differences between conditions. As shown in Fig. 9, the distribution of the scores was mostly positive in both conditions, with only a few negative outliers. Responses from students in the control group resulted in a mean sentiment score of $\bar{x} = 0.451$ ($\tilde{x} = 0.556$, $s_x = 0.400$, $x_{\min} = -0.317$, $x_{\max} = 0.859$), while those in the treatment condition resulted in a mean score of $\bar{x} = 0.460$ ($\tilde{x} = 0.598$, $s_x = 0.331$, $x_{\min} = -0.356$, $x_{\max} = 0.796$).

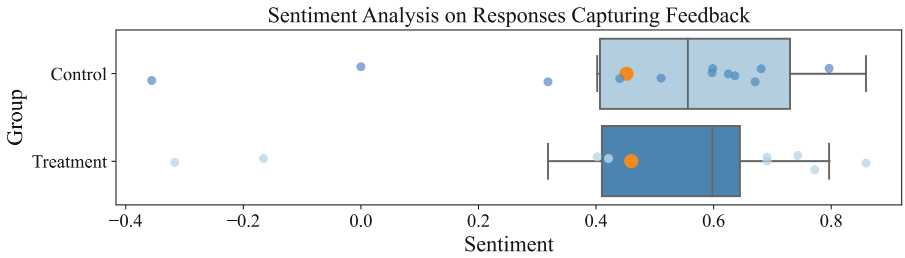


Fig. 9. Feedback

Our qualitative analysis of the first feedback question—which concerned both conditions—showed that all students except two (two treatment) provided a positive comment. Codes such as *good*, *great*, *fun*, *interesting*, and *helpful* were prevalent in students' responses under both conditions. Regarding the negative feedback, while one student provided a minor, off-topic comment regarding the color of the user interface, the other student questioned the need for a chatbot: *"Was the chatbot REALLY necessary? This lesson needs to end on a link to a lesson/tutorial on how to use/configure/install/firststeps/basics/... on ESLint"*.

Nevertheless, four students specifically provided positive feedback regarding their interactions with the chatbot, such as: *"It was helpful and I liked the interactivity with a bot"*. Finally, one student noted that while the interactivity and response time provided by the chatbot were positive, the chatbot would *never* replace the educator: *"ESLint was interesting. It's nice to have an answer directly to our questions. After that, it will never replace the answers of a teacher (even if the answer is direct)"*.

Furthermore, all students in the treatment group ($n = 13$) provided an answer to the second feedback question, which specifically asked about the chatbot and was therefore only visible to students in the treatment group. Although eight students provided positive feedback about the chatbot, four of these comments also included a note about how the chatbot had been *"repetitive"* or *"asked too many questions"*. One student noted the following:

> *"It was nice, but sometimes repetitive. He wished me twice a great day and when I asked how to implement something in the linter, it didn't show me the code. But overall, I find it ludic and it's a nice way to learn since we have interactions. May be interesting for children too. Maybe less with adults."*

Repetitiveness was also observed in the five negative comments, with students urging the chatbot to *"stop asking questions at the end"* or characterizing chatbots as *"pushy salesmen"*. One student provided the following constructive comment:

> *"Sometimes less interaction is more. In this lesson maybe too many interactions are offered and this could be at some point a bit annoying for the user. But still if correctly dosed it may bring some value for the user!"*

5.5 Usability

Both groups rated the usability of the lesson positively. Compared to the UEQ benchmark, in the control group, the results achieved were above average (25% of results better, 50% of results worse) for four dimensions—*attractiveness*, *dependability*, *stimulation*, and *novelty*—while they were good (10% of results better, 75% of results worse) for *efficiency* and excellent (in the range of the 10% best results) for *perspicuity*. In the treatment group, the results achieved were above average for *stimulation*, good for three dimensions—*attractiveness*, *efficiency*, and *dependability*—and excellent for *perspicuity* and *novelty*.

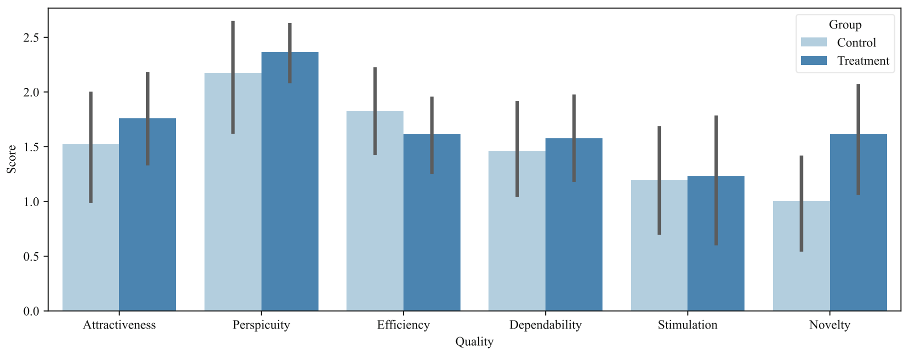


Fig. 10. Results of the User Experience Questionnaire (UEQ)

When comparing between conditions, however, two-sample t-tests did not result in any significant differences across any of the usability dimensions. Nevertheless, it is worth noting that the treatment condition achieved a better usability score in all dimensions, except *efficiency*, and specifically for the *novelty* dimension ($p = 0.0886$). These ratings are summarized in Fig. 10.

5.6 Conversation

For students in the treatment group, our analysis included a sixth aspect regarding the conversations students had with the chatbot. This analysis comprised 150 conversations.

First, all conversations were on topic. Occasionally students opened up the dialogue to cover broader subjects—such as JavaScript in general, instead of JavaScript code style—but these topics still fell within the lesson's scope.

Second—as shown in Fig. 11—approximately two-thirds of the conversations comprised three messages (only one message from the student and two from the

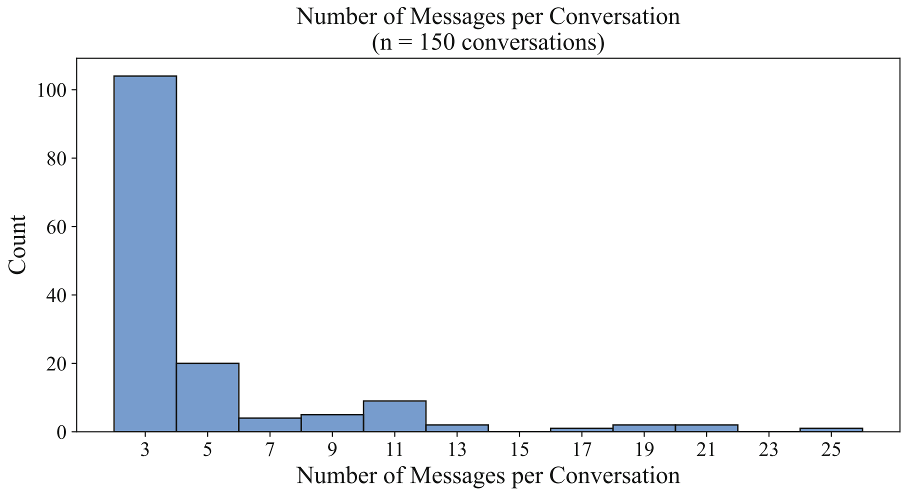


Fig. 11. Number of Messages per Conversation

chatbot). The other two-thirds of conversations were of length greater than or equal to five messages (i.e., at least two messages from the student).

Third, in most conversations, learners had to ask several times for the chatbot to stop asking if they had any further questions (see Fig. 12). Moreover—as mentioned above—in over two-thirds of the conversations, students did not pursue the exchange with the chatbot after one reply (as evidenced by the number of conversations comprising three messages). This means that students ignored the follow-up question from the chatbot, resulting in an unnatural end to the conversation.

Finally, we can divide the cues presented by the chatbot into three *types*: (i) those in which the student is invited to agree or disagree with its statements (*type I*, e.g., *Do you agree?*), (ii) those in which it asks for the student's opinion (*type II*, e.g., *What do you think about... ?*), and (iii) those coming after the exercises, in which it asks students whether they were able to identify a particular set of code style issues (*type III*, e.g., *Did you get all of these issues?*). Out of the 15 cues, seven were of type I, three were of type II, and five were of type III. We analyzed the format of the students' responses for each of these types. These can be short answers (e.g., "*yes*", "*no*", "*OK*") or long answers, i.e., developed and/or justified answers (e.g., "*It can be disturbing at first but it's readable too.*").

As shown in Fig. 13, approximately three-fourths of type I questions result in a short answer, after which most students leave the exchange (reflected in a conversation comprising three messages). On the other hand, we notice that about two-thirds of type II questions lead to long answers. Finally, for type III, we notice that the answer formats are almost equally distributed. By observing the content of these answers, we notice that short answers correspond most of the time to students who correctly identified the issues present in the exercise,

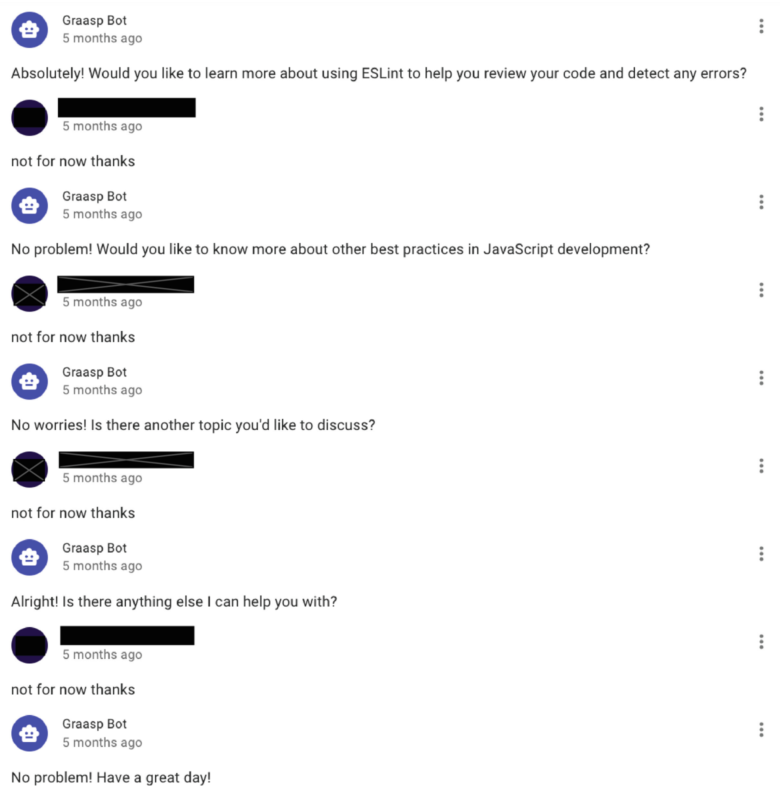


Fig. 12. In many cases, the chatbot kept asking questions even though the learner sought to end the exchange, as shown by the repeated messages above.

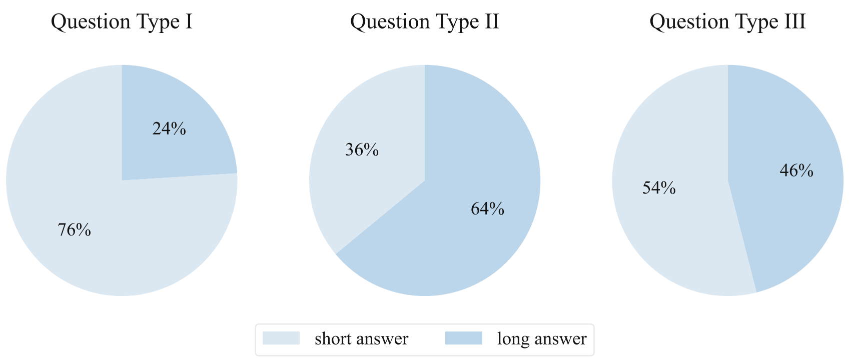


Fig. 13. Length of Answer by Type of Question

while long answers are those of students stating their mistakes and sometimes explaining the reasons for these mistakes or oversights.

6 Discussion

The results of our evaluation did not surface any significant differences between conditions for any of the five aspects considered for both conditions. However, there are a few points that stand out and provide interesting insights into the directions that we can explore in future work.

First, results were more positive in the treatment condition across all five aspects. That is, the mean learning gain was higher, students spent—on average—more time in the lesson, were more positive in reflecting on their performance in the post-test, provided more positive feedback, and rated the lesson higher on the UEQ. Although we emphasize that the differences were not statistically significant, these results offer a positive outlook for the integration of this type of educational chatbot into digital education platforms. At a minimum, these results show that educational chatbots following our architecture can complement pedagogical scenarios without interfering with the learning experience. Extensions of this study with larger cohorts, alternative instruments, and longer exposures could produce more concrete results.

Furthermore, qualitative feedback showed that, in general, learners appreciated the chatbot. Of the 12 students in the treatment condition who provided general feedback through a response to the first feedback question, five explicitly mentioned the chatbot, and only one did so to question whether its integration into the lesson was really necessary. Noting that the chatbot was intended to provide *extra* interaction in the lesson—in order not to compromise the learning experience of students in the control group—the chatbot was indeed designed to not *really* be necessary. However, in answers to the question that specifically asked students in the treatment condition about the chatbot, eight of the 13 students who responded provided a positive comment. The fact that the chatbot was appreciated by the majority of students who were exposed to it is a promising result.

Nevertheless, it is important to acknowledge that the repetitiveness of the questions posed by the chatbot negatively affected the learning experience, as made evident in the qualitative feedback provided by the students in the treatment group. These repeated questions were a result of the chatbot's configuration. More specifically, the chatbot's prompt (see Sect. 4.1) included the following instructions: *After each response, the chatbot gives the student one or two options to continue the conversation.* Although this strategy was effective in engaging students, it failed to capture the moment when the student wanted to stop interacting with the chatbot. Thus, this strategy quickly backfired and resulted in some students referring to the chatbot as *pushy* and *annoying*. This result sheds light on how what may appear to be a minor configuration detail could potentially have a negative impact on the user experience when using powerful LLMs.

Third, a close inspection of Fig. 7 shows that the average time spent in the phases that included chatbots (Styling, Best Practices, Solutions, and Chatbots) was actually longer for the control condition than for the treatment condition. While these differences are not significant, the consistency of these results across the four phases stands out. We would have expected students in the treatment condition to spend more time in these phases due to the extra interaction with the chatbot that students in the control condition—who only had to read the accompanying text—did not have to engage with. However, it could be the case that students in the treatment condition favored focusing on the explanation provided by the chatbot, which served as an interactive summary of what was contained in the text. Hence, the chatbot might have provided a faster way of learning the issue that was captured in each code snippet or simply a way to guide a student's focus through the exercise.

It is also worth pointing out that the results of the UEQ show that the differences between conditions were most significant for the *novelty* dimension. For this dimension, the ratings provided by students in the treatment condition are significantly higher than those provided by students in the control condition at the $p < 0.1$ level. The need for educational technologies to remain novel and attractive is particularly relevant in light of the rapidly changing technological landscape. Learning technologies and digital education platforms that achieve positive usability results in these dimensions are likely to have an advantage in attracting learners and keeping learners engaged.

With respect to the exchanges held between students in the treatment group and the chatbot, a first positive result is that conversations were all on topic. The fact that about a third of conversations consisted of five or more messages is an indicator that students took advantage of the opportunity to interact with the chatbot. However, conversations often lacked naturalness. As discussed above, this was caused by the fact that the chatbot was configured to end its messages with a question that was supposed to encourage the student to reply. This caused problems when the student wanted to end the conversation. Finally, the formulation of the cue appeared to have had an effect on whether student responses were short or long. While type I cues asking the student just to agree or disagree did not lead to long answers, type II cues asking the student to reflect were better at encouraging students to provide more in-depth answers.

7 Conclusion

In this paper, we presented the design of an architecture aimed at supporting developers in education in integrating LLM-powered chatbots into digital education platforms. We then conducted a case study comprising a between-subjects controlled experiment with 26 software engineering students. Half of the students completed the lesson with support from Graasp Bot—an educational chatbot embedded in an application implemented following our architecture—while the other half completed the lesson without support from the chatbot. Although there were no significant differences across the aspects considered in our evaluation, the results of our study can help optimize our prompt engineering strategy and provide useful examples for researchers and educators looking to incorporate LLM-powered chatbots into their practice. Furthermore, given that learning gains were not impacted by the presence of the chatbot, these findings reinforce the idea that educational chatbots could serve to provide additional information when educators are not available.

It is also important to note that there are a number of limitations that could have affected our findings. First, while our sample size was appropriate for our mixed methods experiment, expanding our study to include more subjects could help in the detection of differences between conditions, especially in pedagogical scenarios with shorter durations, where differences could be more subtle than expected. Similarly, increasing exposure by conducting semester-long or longitudinal studies could also serve to better identify the differences that emerge between conditions. Second, exploring this interaction strategy with different pedagogical scenarios and subject matter could help generalize the applicability of our architecture. Finally, incorporating standardized instruments to measure learning gains, self-reflection, and engagement could help reveal more interpretable results regarding how LLM-powered chatbots can provide support in educational contexts.

References

1. Airbnb: Airbnb JavaScript Style Guide (2012). https://airbnb.io/javascript/
2. Bach, S.H., et al.: PromptSource: An Integrated Development Environment and Repository for Natural Language Prompts (2022). https://doi.org/10.48550/arXiv.2202.01279
3. Bergin, J.: Fourteen pedagogical patterns. In: Devos, M., Rüping, A. (eds.) Proceedings of the 5th European Conference on Pattern Languages of Programs (EuroPLoP 2000). Universitaetsverlag Konstanz, Irsee, Germany (2000)
4. Charmaz, K.: Constructing Grounded Theory: A Practical Guide through Qualitative Analysis. Sage, London (2006)
5. Farah, J.C., Spaenlehauer, B., Rodríguez-Triana, M.J., Ingram, S., Gillet, D.: Toward code review notebooks. In: 2022 International Conference on Advanced Learning Technologies (ICALT), New York, NY, USA, pp. 209–211. IEEE (2022). https://doi.org/10.1109/ICALT55010.2022.00068
6. Gillet, D., Vonèche-Cardia, I., Farah, J.C., Phan Hoang, K.L., Rodríguez-Triana, M.J.: Integrated model for comprehensive digital education platforms. In: 2022 IEEE Global Engineering Education Conference (EDUCON), New York, NY, USA, pp. 1586–1592. IEEE (2022). https://doi.org/10.1109/EDUCON52537.2022.9766795
7. Hunter, J.D.: Matplotlib: a 2D graphics environment. Comput. Sci. Eng. **9**(3), 90–95 (2007)
8. Hutto, C.J., Gilbert, E.: VADER: a parsimonious rule-based model for sentiment analysis of social media text. In: Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media, Ann Arbor, MI, USA, pp. 216–225. AAAI (2014). https://doi.org/10.1609/icwsm.v8i1.14550
9. Jiang, Z., Xu, F.F., Araki, J., Neubig, G.: How Can We Know What Language Models Know? (2020). https://doi.org/10.48550/arXiv.1911.12543

10. Johnson, S.C.: Lint, A C Program Checker. Technical report, Bell Laboratories, Murray Hill, NJ, USA (1978)
11. Kasneci, E., et al.: ChatGPT for Good? On Opportunities and Challenges of Large Language Models for Education (2023). https://doi.org/10.35542/osf.io/5er8f
12. Laugwitz, B., Held, T., Schrepp, M.: Construction and evaluation of a user experience questionnaire. In: Holzinger, A. (ed.) USAB 2008. LNCS, vol. 5298, pp. 63–76. Springer, Heidelberg (2008). https://doi.org/10.1007/978-3-540-89350-9_6
13. OpenAI: Introducing ChatGPT (2022). https://openai.com/blog/chatgpt
14. Reynolds, L., McDonell, K.: Prompt Programming for Large Language Models: Beyond the Few-Shot Paradigm (2021). https://doi.org/10.48550/arXiv.2102.07350
15. Song, D., Oh, E.Y., Rice, M.: Interacting with a conversational agent system for educational purposes in online courses. In: 2017 10th International Conference on Human System Interactions (HSI), Ulsan, South Korea, pp. 78–82. IEEE (2017). https://doi.org/10.1109/HSI.2017.8005002
16. Tómasdóttir, K.F., Aniche, M., van Deursen, A.: The adoption of JavaScript linters in practice: a case study on ESLint. IEEE Trans. Software Eng. **46**(8), 863–891 (2020). https://doi.org/10.1109/TSE.2018.2871058
17. Zakas, N.C.: ESLint (2013). https://eslint.org/
18. Zhao, T.Z., Wallace, E., Feng, S., Klein, D., Singh, S.: Calibrate Before Use: Improving Few-Shot Performance of Language Models (2021). https://doi.org/10.48550/arXiv.2102.09690

Motivating Learners with Gamified Chatbot-Assisted Learning Activities

Alexander Tobias Neumann[1](✉), Aaron David Conrardy[2](✉), Stefan Decker[1], and Matthias Jarke[1]

[1] RWTH Aachen University, Aachen, Germany
{Neumann,Decker,Jarke}@dbis.rwth-aachen-de
[2] Luxembourg Institute of Science and Technology, Esch-sur-Alzette, Luxembourg
Aaron.Conrardy@list.lu

Abstract. Chatbots belong to the many technological novelties that could be employed in educational settings. Already being shown successfully in contexts such as being able to answer frequently asked questions, more advanced educational chatbots have been developed to support learners in their studies. However, despite their potential, chatbots have faced challenges related to low interaction rates, often attributed to factors such as insufficient user engagement, limited conversational capabilities, and the inability to sustain learners' interest over time. To address this issue, we explore the concept of gamification, a strategy frequently employed to boost extrinsic motivation in existing activities or applications. By incorporating game-like elements into an application, students may be more inclined to engage with it. In this work, we present the design, implementation, and evaluation of a gamified Quizbot called Quiz-GBot tailored for educational purposes. The chatbot records student interactions in a learning record store and subsequently gamifies these learning activities to stimulate student engagement. Our evaluation with 54 participants shows the support for tailored gamification in educational chatbots, particularly among users with gaming experience. Additionally, gamification elements impacted participants differently depending on player type, revealing the necessity for personalization to further improve user enjoyment. Importantly, features such as leaderboards can enhance motivation but should be personalizable.

Keywords: Chatbots · Gamification · Learning Activities

1 Introduction

E-Learning is described as learning driven by digital electronic tools and media [11]. The fast and constant development of technology requires educational institutions to quickly adapt to these changes and include them in education [9].

Gamification and chatbots have emerged as promising tools for enhancing learners' motivation and engagement in educational contexts [4,10]. Gamification refers to the use of game design elements, such as points, badges, and leaderboards, in non-game contexts to make them more enjoyable and motivating [7].

H. Xie et al. (Eds.): ICWL 2023, LNCS 14409, pp. 189–203, 2023.
https://doi.org/10.1007/978-981-99-8385-8_15

These game elements are supposedly part of the reason why games are fun, thus including these in non-game-contexts could be seen as adding motivational elements to existing applications.

Chatbots, on the other hand, are computer programs that simulate human conversations and can interact with learners in natural language. They can answer basic questions [10], send reminders to students and notify them about new learning content [2] or assess the students based on learned content [16].

We claim that the combination of gamification and chatbots in educational settings, referred to as gamified chatbot-assisted learning activities, has the potential to provide personalized and interactive learning experiences that can increase learners' motivation and improve their learning outcomes. Despite the growing interest in gamified chatbot-assisted learning activities, there is still a need for empirical evidence to underpin their effectiveness and to understand how they can be designed to optimize learners' motivation and engagement. Therefore, this paper aims to investigate the potential of integrating gamification into chatbots to enhance user motivation and engagement in learning activities while also investigating how some demographic attributes might affect the results. To address this objective, we aimed to gamify a Quizbot by making use of standardized data and formulated the following research questions:

RQ1: ***How can gamification be integrated into educational chatbots to support Technology Enhanced Learning (TEL)?*** This involves investigating which technical methods and tools can be used to effectively integrate gamification elements into a chatbot system, how these integrations should be represented and how they can support specific learning outcomes and objectives.

RQ2: ***Could gamification positively affect the motivation to interact with chatbots?*** To this end, we investigate whether users feel interested and invested in communicating with a gamified chatbot over a short evaluation. We additionally investigate whether motivation is affected by previous game experience (RQ2.1) and player types (RQ2.2).

RQ3: ***How high is the acceptance in regards to a gamified chatbot?*** Here, we explore user acceptance of a gamified chatbot with integrated elements. We additionally investigate whether acceptance is affected by previous game experience (RQ3.1) and player types (RQ3.2).

2 Background and Related Work

The background and related work section presents theoretical concepts used during the implementation and an overview of the existing literature on gamification and chatbots in various contexts, highlighting the different gamification elements used to motivate users to engage with the application.

Aspects of Motivation. Motivation, which is defined as "[...] *being moved to do something* [...]" [18], is seen as directly linked to learning success [3]. The lack thereof is seen as one of the main reasons for students not engaging regularly with learning content. For that purpose, motivational models have been designed to explain which aspects of a person or a system affect motivation, which can then be used to design learning content or learning systems to aim at fulfilling the different aspects [6,14]. Ryan and Deci's Self-Determination Theory (SDT) posits that the three innate psychological needs *Autonomy, Competence and Relatedness* need to be fulfilled to foster motivation [6]. Similarly, Keller's ARCS motivational model describes the four components *Attention, Relevance, Confidence and Satisfaction* as the human conditions needed to be met to foster motivation [14]. Research has shown that gamification has the potential to foster the user's motivation, which implies that gamification elements fulfill the different components of the motivational models [21].

Gamified Chatbots. *CiboPoliBot* is a gamified Telegram[1] chatbot that aims to teach healthy lifestyles to children aged 8–14 [8]. The design process involved using the HEXAD gamification framework to select gamification elements that fit the player types based on learning goals, tasks, and target groups [15]. Then, Fadhil and Villafiorita integrated leaderboards and points into the bot [8]. The bot operates as an interface for a game, where users collect virtual foods, give or throw food away, or eat the food based on the result of a dice roll. Points are rewarded based on the collection of choices, and players are placed on a leaderboard at the end. However, the bot's effectiveness has not been evaluated, it is currently impossible to conclude its motivational outcomes. Additionally, there is a lack of examples of how points and leaderboards would be implemented in a chat environment.

Escapeling is a gamified Telegram bot for English learning in secondary schools where English is a second language for students [12]. It promotes collaborative learning through three types of English tasks, where learners discuss and solve problems together. The bot engages users with a narrative, portraying them as trapped on an alien ship and requiring English problem-solving skills to escape. In addition to the narrative, the chatbot includes badges, streaks, unlockable content, and displayable achievements. User achievements are displayed in private chats, and new achievements are announced in group chats to activate the innate social comparison mechanism and further motivate users. Two evaluations of the app's usability, user satisfaction, and success were conducted, indicating high performance and satisfaction scores. However, the evaluations did not emphasize the motivational aspects of the gamification elements, and no visualizations of the gamification elements were provided.

The popular live streaming platform *Twitch.tv* has gamified its user interface and chat environment for streamers and viewers [13,20]. In addition to Twitch.tv's built-in gamification, streamers were previously able to import third-

[1] https://telegram.org/.

party chatbots such as Streamlabs[2], DeepBot[3] and PhantomBot[4], which offer similar gamification features. Viewers receive points for commenting, following, subscribing and donating, which can be used to unlock special emotes or take part in mini-games that reward more points if won. These bots feature a leveling system, leaderboards, and customizable elements like badges, progress bars, and a streak system, using commands instead of Natural Language Processing (NLP). Despite Twitch.tv's popularity, there has been no research on the effects of gamification elements or bots on user interaction or viewership.

Discord, a communication platform that offers private and public chat channels, has gamified chatbots like *Tatsu*[5] and *Arcane*[6] that can be added to any server, allowing users to earn currency points by sending text messages that can be used to purchase roles, badges, or virtual items. Users earn experience points by completing daily quests and leveling up to unlock special server features in both Tatsu and Arcane. Tatsu offers visualizations of progress, a user profile card, and a leaderboard upon request, while Arcane includes an experience point system, leaderboards, and moderator actions.

Table 1 contains a comparison of the previously discussed gamified bots and our proposed system. We noticed that existing gamified chatbots propose gamification elements with interesting designs. However, many of them have not been tested in an educational setting, lack empirical evaluations to support their effectiveness, are not open source, or use non-standardized data. These limitations restrict their potential for gamifying arbitrary chatbot data and reusability. Our solution aims to address and overcome these issues.

Table 1. Comparison of gamified chatbots and frameworks.

	Points	Levels	Leaderboards	Badges	Streaks	Achievements	NLU	Open source	Standardized Data Source
CiboPoliBot	✓	–	✓	–	–	–	–	–	–
Escapeling	–	–	–	✓	✓	✓	–	✓	–
Twitch.tv bots	✓	✓	✓	✓*	✓*	–	✓	–	–
Discord bots	✓	✓	✓	✓	–	✓	–	✓	–
Our solution	✓	✓	✓	✓	✓	✓	✓	✓	✓

* are features that need additional implementation.

3 Concept and Realization

In this chapter, we discuss the implementation of our framework to enable gamification in chatbots, covering the technologies used, the integration of game-like elements and the creation of Quiz-GBot, a gamified Quiz-Bot.

[2] https://streamlabs.com/chatbot.
[3] https://deepbot.deep.sg/.
[4] https://phantombot.tv/.
[5] https://tatsu.gg/.
[6] https://arcane.bot/.

Firstly, we decided to gamify standardized interactions stored in a Learning Record Store (LRS)[7] database in xAPI[8] format. In contrast to the previously mentioned gamified chatbots implemented for specific use cases, our approach allows the gamification of arbitrary interaction data using rules based on the content of the data. Regarding the gamification elements, we adopted a design-based research approach to identify and implement suitable gamification elements in the Quiz-GBot prototype. Initially, we pinpointed areas where gamification could enhance user engagement. Using these insights, we brainstormed, designed, and iteratively tested various gamification features, refining them based on user feedback. This systematic, user-centric, and iterative process ensured the gamification elements were not just added for novelty but genuinely enhanced the Quiz-GBot user experience. Some elements, such as achievements and level notifications, were displayed as text, while others like the profile card and the achievement list were integrated as images. We stick to these two basic formats, instead of opting for platform-dependent features, as it ensures cross-platform compatibility and accessibility. The use of text and images still manages to accommodate users with diverse devices and screen sizes while still offering a visually engaging and immersive user experience while maintaining compatibility across all platforms. While platforms like Telegram, Discord[9], and Slack[10] have more polished user interfaces (buttons, checkboxes, select menus, . . .), and hence could potentially offer a better user experience, they were not chosen because they are not open-source. This could lead to privacy concerns when using these platforms for learning activities. Thus, our design decisions were driven by a balance between user experience, privacy considerations, and the importance of broad accessibility. To further reason our choices for the implemented gamification elements, we created a mapping (see Table 2) of gamification elements to motivational aspects.

Table 2. Mapping of Gamification Elements to Motivational Aspects

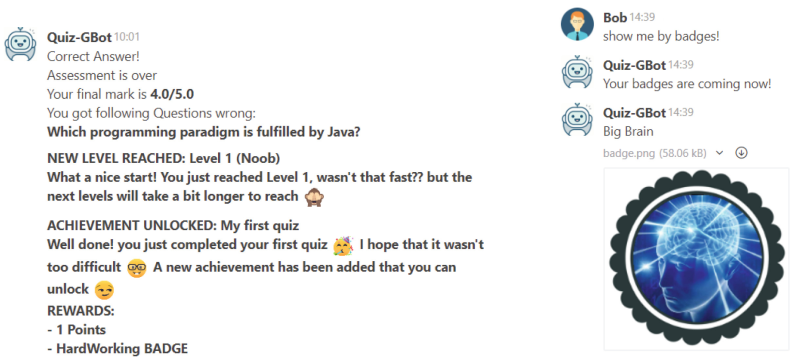
	Autonomy	Competence	Relatedness	Attention	Relevance	Confidence	Satisfaction
Points	–	✓	–	–	–	✓	✓
Levels	–	✓	–			✓	✓
Leaderboards	–	✓	✓	–	–	✓	✓
Achievements	✓	✓	–	✓	✓	✓	✓
Badges	–	–	–	✓	✓	✓	✓
Streaks	✓	–	–	✓	–	–	✓

This mapping showcases which gamification element we expect will affect the different motivational aspects to then foster motivation. Note that, it is not

[7] https://xapi.com/learning-record-store/.
[8] https://xapi.com/.
[9] https://discord.com/.
[10] https://slack.com/.

guaranteed that these aspects will be affected, as individual users might perceive them differently and the implementation and chosen semantics also play a role.

Points, Levels, Badges and Achievements. Our Gamification Framework enables earning points through actions or achievements, unlocking other gamification elements. When defining levels, the game creator must specify attributes such as the level number, name, necessary points to reach the level, and the notification message. Rules specify the available system actions, which assist users in progressing in the gamified environment and may also indicate the number of points users earn upon action completion. We utilize xAPI statements stored in an LRS to track users' activities effectively. The xAPI statements offer granular details on each user's interactions, making them invaluable for monitoring their journey and accomplishments within the game environment. We maintain a structured database record that maps each activity to its corresponding action, including the points it yields. This database not only aids in accurately awarding points but also provides a clear benchmark for the number of points necessary to transition to the subsequent level. A user will automatically advance to the next level once they have accumulated the necessary points. An example is shown in Fig. 1a. For badges, the game creator simply uploads a picture that will be resized by our Gamification Framework, and provides a badge name and ID. For achievements, in addition to the basic attributes such as achievement name, ID, and notification, the game creator must describe the achievement and the rewards that a user obtains upon unlocking it. These rewards can include a point value and/or a badge that will be awarded to the user. Figure 1b shows how a user asks for his badges.



(a) Progressing and unlocking achievements. (b) Requesting badges.

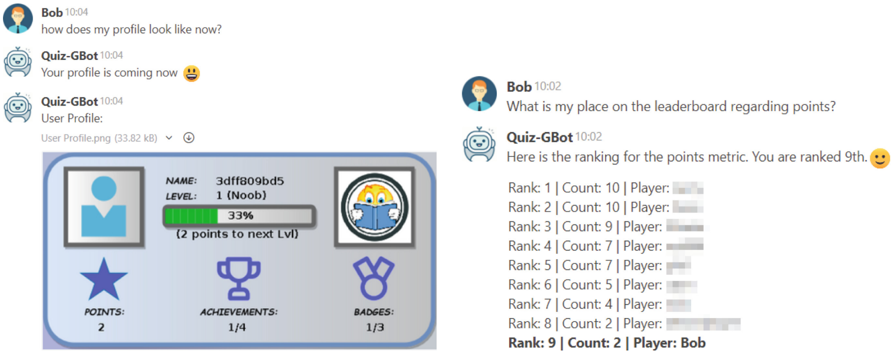
Fig. 1. Screenshots of the gamified Quiz-GBot showing a user: (a) reaching a new level and unlocking achievements, and (b) requesting his earned badge.

Quests and Streaks. The quest element in our Gamification Framework consists of an unlockable achievement and the required actions to complete the quest and unlock the achievement. The game creator can choose which achievement to unlock from the list of existing achievements and which action to perform, along with the number of occurrences of the chosen action to unlock the achievement. Similar to quests, streaks allow the game creator to choose which action causes the streak level to increase. Moreover, achievements that can be unlocked have to be specified for streaks, along with the streak level at which they are unlocked. Streaks also require a duration value to specify how much time should pass between two consecutive action triggers until a streak level is lost. If a streak is lost, the level falls back to level zero.

Player Profile and Leaderboards. Our Gamification Framework provides a player profile card to reflect the current progress of users, similar to the Tatsu Discord bot. To avoid overwhelming users with information, the displayed information on the player profile card was kept minimalistic. The information displayed includes the collected points, current level, required points to reach the next level, number of unlocked and missing achievements, and number of unlocked and missing badges. To add a light form of personalization, the player's nickname and a badge earned can be added to the profile card. The profile card was implemented as a PNG file that is sent to the user, as this option results in a visually appealing result that is not exclusive to any chat platform. The final design of the template contains an extra placeholder square, which can be used to personalize the profile card by inserting badges at the reserved emplacement and can be personalized by the user. A personalized and completed profile card example is shown in Fig. 2a. Users can also request a leaderboard that displays users' scores based on a selected activity/metric and helps them compare their progress to other users. The leaderboards can be based on metrics that are present in every application gamified by our Gamification Framework, these metrics being collected points, unlocked achievements and collected badges. Additionally, they can also be based on possible actions in the gamified application, thus being application-specific. On the chat platform, the leaderboard is displayed as a plain text message using minimal formatting containing the three columns rank, number of occurrences/score and the member ID, which is a pseudonym to keep the users anonymous. Figure 2b shows an example leaderboard that displays scores based on points.

The entire framework and the Quiz-GBot implementation are open-source and available on GitHub[11]. The gamification component for Quiz-GBot was developed using Java and serves as a wrapper service for our Gamification Framework employed in this project. Its primary purpose is to connect chatbots with gamification elements, enabling seamless integration and communication between the two by gamifying interaction data stored in a LRS database.

[11] https://github.com/rwth-acis/Gamification-Framework.



(a) User asking for his player profile. (b) Asking for leaderboard.

Fig. 2. Screenshots of the gamified Quiz-GBot showing a user requesting: (a) his player profile, and (b) the leaderboard.

4 Evaluation

The evaluation section primarily investigates the effects of a gamified chatbot on motivation, acceptance, and usability. By examining these factors, we aim to determine the feasibility and effectiveness of gamifying chatbots in educational contexts. We additionally compare the results based on the users' previous game experience and player types, which can provide further insights into the findings.

4.1 Design, Procedure and Tasks

The evaluation consisted of two surveys and an actual testing session. An initial survey captured demographics, while a second focused on feedback regarding the tested project. The participants were invited to individual online meetings, which were held over Zoom. The recruitment process involved posting the study invitation in forums and sending individual invitations to acquaintances, which ultimately resulted in 54 participants participating in the evaluation. The evaluation commenced with an introduction to gamification and the study's objective. The participants were provided with a link to the non-gamified version of the Quiz-Bot chatbot. They were prompted to greet the bot and initiate a quiz, which was offered on a variety of topics, including science, history, and pop culture. While participants engaged with various topics, analyzing the correlation between their selections and their educational background or interests was not within the scope of our research. Upon completing the quiz, the participants proceeded to the gamified version of Quiz-Bot, Quiz-GBot, where they first had to request gamification elements using methods such as selecting options, asking with text, or using a command. The participants were prompted to request the profile card and unlock any achievement. After completing a quiz, participants received notifications of unlocking a new level and achievement and should check their progress on the leaderboard and their profile card. The study concluded with the second survey, which sought their feedback on the experience.

(a) Previous Game Experience ($n = 54$).

(b) Identified HEXAD player types.

Fig. 3. Game experience and identified HEXAD player types among the participants.

4.2 Results and Comparison

The first survey aimed to gather participants' demographic information and their previous gaming experience. Out of the 54 participants, 39 were male and 15 were female, with 44 participants falling into the age range of 19–25 and ten participants being between the ages of 26–35. Figure 3a shows a breakdown of participants' previous gaming experience, which was categorized into three groups: those with considerable gaming experience (39 participants), those who were somewhat acquainted with games (12 participants), and those who claimed to have almost no gaming experience (3 participants). Furthermore, the Player Type questionnaire from the HEXAD gamification framework was also included in the first survey to categorize participants into different player types [15]. The results of the categorization can be seen in Fig. 3b. Note that the numbers don't add up to 54, as some people can be categorized as multiple player types. Following the evaluation session, we administered a survey based on existing motivational models, Technology Acceptance Model (TAM), and questions directly comparing the gamified and non-gamified versions of the bots [5,17].

Table 3. Participants' responses on a five-point Likert scale (ranging from 1 = strongly disagree to 5 = strongly agree) to questions related to motivational models ($n = 54$).

Aspect	Question	$\bar{x}(\sigma_x)$
A1	I felt interested in communicating with the gamified bot	4.37(±0.62)
	Communicating with the gamified bot felt like a drag	2.02(±1.03)
A2	The gamified information in the gamified bot kept my attention	4.19(±0.80)
	The gamified elements made me want to continue interacting with the bot	4.26(±0.76)
	I found the gamified elements distracting	1.61(±0.83)
R1	The used elements were known to me	4.15(±0.98)
	I didn't face difficulties understanding the gamified elements and how they work	4.64(±1.12)
C1	It was easy to understand why I unlocked achievements	4.65(±0.62)
	I always understood when I unlocked an achievement/reached a new level	4.56(±0.74)
	I felt that the profile card reflected my progress fittingly	4.48(±0.57)
S1	Unlocking achievements and receiving rewards with the gamified bot made me feel rewarded for my effort	4.30(±0.86)
R2	My place on the leaderboard made me feel better	3.54(±1.08)
	I enjoyed seeing myself on the leaderboard next to others	3.93(±0.97)

Questions Based on Motivational Models. The questions were adapted versions of those used in previous studies, such as the evaluation of gamified E-Learning systems by Shi and Cristea [19] which is based on Self-Determination Theory by Ryan and Deci, and the ARCS model of motivation by Alcasoda and Balaoro [1]. The questions were designed to answer the research questions related to the effects of gamification on motivation. The questions focused on aspects of the motivational models, such as "Autonomy (A1)", "Attention (A2)", "Competence (C1)", "Relevance (R1)", "Relatedness (R2)" and "Satisfaction (S1)". Table 3 shows the mean and standard deviation values of participants' responses to the motivational models' questions.

Questions Based on Technology Acceptance Model. The second part of the survey aimed to answer the research questions regarding technology acceptance. We used questions that were inspired by a study evaluating an e-learning system for university students [17] with questions based on the TAM proposed by Davis et al. [5]. We adapted these questions to our project while ensuring that they addressed the different constructs that make up the TAM. The results are presented in Table 4, which shows the mean and standard deviation of users' responses to questions regarding the perceived ease of use, usefulness, attitude, behavioral intention, and self-efficacy of the gamification elements in the chatbot-based learning system.

Table 4. Users' perceptions of gamified chatbot-based learning activities on a seven-point Likert scale ranging from 1 = strongly disagree to 7 = strongly agree ($n = 54$).

Aspect	Question	$\bar{x}(\sigma_x)$
PE	I find the gamification elements easy to use	5.98 (±1.17)
	Learning to use the gamification elements was easy for me	6.17 (±1.02)
	It was easy to become skillful at using the gamification elements	6.13 (±0.85)
PU	The gamification would improve my learning performance (how well I absorb learning content and can apply it)	5.55 (±1.33)
	The gamification would improve my academic productivity (studying/working time)	5.74 (±1.08)
AT	Studying through a gamified bot is a good idea	5.92 (±0.83)
	I am positive toward a gamified bot	6.19 (±0.68)
	I feel that the gamified elements make the bot more enjoyable	6.40 (±1.02)
BI	I would turn on the gamified bot elements if given a choice	6.09 (±1.20)
	I intend to use the gamified bot elements a lot given the opportunity	5.72 (±1.31)
	I would interact with the chatbot more due to the gamification	5.90 (±1.32)
SE	I felt confident using the gamified elements in the system	6.04 (±1.08)
	I have the necessary skills to use gamified elements	6.30 (±1.06)

Questions Comparing Gamified and Non-gamified Chatbot. This last block of questions focuses on comparing both the non-gamified and gamified versions of the chatbot. Out of the 54 participants, 45 preferred the gamified chatbot, 8 saw both gamified and non-gamified versions as equally good and 1 participant preferred the non-gamified version. The participants also could explain their chatbot choice in an open text field. Some found the gamified version more interesting but more time-consuming, which could be a drawback during exam preparation. Others mentioned that the gamified bot felt "unserious" or distracting at times. Yet, regarding the positive aspects, participants mentioned that the gamified version felt "Fun" and more interesting than the non-gamified version which was considered boring by some. Achievements would give purpose to the interactions and let them feel in control of what they want to achieve. The visualizations of progress and immediate feedback on unlocking achievements or reaching a new level were also praised. Some mentioned that the gamified version compensated for the lack of incentives of the non-gamified version. The final two statements directly compared both chatbots in terms of fun and immersion. The high score for the statement "Interacting with the gamified bot was more fun than with the simple bot" (AVG = 4.60; SD = 0.49) confirms that gamification indeed made the bot more to interact with as opposed to the non-gamified version. While fun is not necessarily an indication of a better learning experience, the users' enjoyment increased, which could potentially increase the motivation to interact with the bot. In terms of immersion, the statement "I felt more absorbed while interacting with the gamified bot than with the simple bot" (AVG = 4.30; SD = 0.66) also ended up with a high average, indicating that generally, participants were more focused on the interaction with the gamified bot than with the simple version. This hints towards the notion that gamification makes the participants concentrate more on the interaction.

Comparison Based on Game Experience. We investigated the impact of previous game experiences on the user's perception of a gamified chatbot. The results showed that participants who never played games rated the gamified bot with a lower score than the other two groups, suggesting that those who do not play video games may not enjoy a gamified chatbot as much. However, the low number of participants in this group limits the certainty of this conclusion. Interestingly, participants in this group did not feel rewarded for their efforts when receiving achievements and expressed opposition towards the leaderboard. Nevertheless, they still agreed that the gamified bot was more fun and immersive than the simple bot, indicating that gamification can still improve the user experience for non-gamers. When comparing the two other groups, the "Rarely" group answered similarly to the "Regularly" group. Although the difference was often not significant, it still suggests that the amount of game experience does not necessarily correlate with a better perception of gamification. However, the "Regularly" group had a much higher participant number, which may have affected the results. Furthermore, the vagueness of the categorization of "Rarely" and "Regularly" may have led some participants who would rather fit in the more experienced category to classify themselves as more inexperienced participants. Overall, it can be concluded that people with some form of gaming experience are more likely to enjoy gamified elements in a chatbot, which may increase their engagement with the application. Our findings suggest that the gamified bot was well-received among all three groups, indicating that gamification can be a valuable tool to enhance the user experience of chatbots, regardless of the user's level of game experience. These findings are supported by the results of the bot preference, which showed that the majority of participants in both the "Rarely" and "Regularly" groups preferred the gamified bot over the simple bot, further emphasizing the positive impact of gamification.

Comparison Based on Player Type. At first glance, there do not seem to be any notable differences in the participants' answers based on their player type. Regarding how the different results match the expected behavior of the HEXAD gamification framework player types, we found some interesting results. The "Socializers", defined by their need to interact with others, answered most negatively when asked whether they felt annoyed during the bot conversation, as these types of players prefer interaction with real people instead of a computer program. "Achievers", who focus on mastery and learning to improve themselves, generally felt that learning to use the gamified elements was not difficult and claimed they became skillful at using them. Regarding unlocking achievements, the "Player" type seemed to enjoy them the most, which is fitting as this group is defined as focusing on rewards the most. "Philanthropists" scored the achievements the lowest out of all groups, which again fits as this group is described as not caring about rewards. "Free Spirits" seemed to be the most skeptical about leaderboards, which potentially means that leaderboards should be removed for this group.

In its current state, the gamified bot mainly consists of reward-based elements such as achievements and badges. This fact might be tied to the result that the "Player" users answered the most positively when questioned about their intention of use outside of the evaluation. However, the leaderboard was received negatively by some players but also positively by some. To address this, adding a new function that hides a player from the leaderboard might be reasonable. Overall, all player types seemed to enjoy the gamified elements during the short evaluation, which was focused on gathering a general first impression of the gamified chatbot. However, a more in-depth study will be needed to focus on individual elements and how they affect users, while also collecting user data to see which elements were used more.

4.3 Discussion

The present study aimed to investigate the impact of gamification on the motivation and acceptance of chatbots. Firstly, we aimed to answer RQ1 with our implementation of a gamified chatbot. The positive results suggest that we managed to create a somehow successful prototype. Yet, there still is room for improvement, as suggested by participants' comments on different aspects such as better visualizations, new features and more. To answer RQ2, we analyzed the results of the survey questions related to the motivational aspects, behavioral intention and bot preference. Overall, the results suggest that gamification might increase the motivation and willingness to interact with a chatbot. Although some outliers showed less interest, they did not view gamification negatively but rather believed it would not increase their motivation. We further investigated subquestions of RQ2, namely RQ2.1 and RQ2.2. The results of RQ2.1 suggest that participants who previously played video games enjoyed the gamified bot more than those who never played. This implies a correlation between the previous game experience and motivation to interact with the gamified bot. In contrast, we did not find any significant correlation between player types and motivation in RQ2.2. Moving on to RQ3, our results suggest that the acceptance of the gamified bot is positive. As seen in the discussion of the results, the different constructs of the used TAM ended up with averages in the "I agree" category. The positive outcome of questions about usability and perceived usability further confirms our belief that the gamified elements made a good impression on the participants. Regarding subquestions of RQ3, we investigated the correlation between acceptance and previous game experience (RQ3.1). The results suggest that people who never played games before tended to be less in favor of the bot than those who did. In terms of the correlation between player types and acceptance (RQ3.2), we did not find any noteworthy results. The results suggest that each player type would enjoy and accept the bot similarly. Overall, our findings suggest that gamification can increase the motivation and acceptance of chatbots. Previous game experience seems to be a significant factor affecting motivation and acceptance of gamified chatbots. The results highlight the importance of the preferences of participants when designing gamified chatbots.

Threats to Validity. Threats to the validity of our evaluation results should be acknowledged to ensure that the conclusions made from this study are not misinterpreted or overstated. One potential limitation of our study is the different sampling sizes of the analyzed categories. In particular, the groups of participants who mentioned having no game experience and the "Disruptor" player type had smaller sample sizes, which may have resulted in less reliable data. Moreover, given that motivation and acceptance are complex constructs, it is plausible that our study, conducted over a relatively short period, may have not captured all the relevant aspects of these constructs. Thus, to address these limitations, we plan to conduct a longer study that will collect usage data and survey responses over a more extended period, which we hope will provide a more comprehensive picture of the effects of gamification on chatbot interaction, enabling us to draw more robust conclusions.

5 Conclusion, Implications and Future Work

In conclusion, this paper has presented a novel approach to increase learners' motivation to interact with educational chatbots by integrating gamification elements. We identified and explored three distinct research inquiries on combining gamification and chatbot technology: how gamification and chatbots may be effectively combined to promote user engagement, the extent to which gamification serves to enhance motivation to interact with chatbots and the users' acceptance of a gamified chatbot and its elements. We proposed a framework to gamify chatbots, independent of the chat platform, using standardized interaction data stored in a LRS and created Quiz-GBot, a gamified Quiz-Bot. The use of the standardized data format allows for the easy gamification of any educational chatbot, opening possibilities to include gamified chatbots in different TEL scenarios. The results of the evaluation of Quiz-GBot, which we based on different motivational models, showed that the gamification of educational chatbots could lead to increased user motivation and acceptance. Participants expressed enjoyment and satisfaction with the different gamification elements and reported an increased motivation to interact with the chatbot. However, our findings also suggest that while gamification can be an effective method for increasing user motivation, it may not be suitable for all users depending on prior gaming experience and other attributes. Yet, we saw that when categorizing users using the HEXAD framework, it made correct predictions about the users' preferences. This leads to the notion that the HEXAD framework can be used as a base for implementing personalization into gamified bots. To improve the gamified bot, we plan to address minor technical issues, such as improving message formatting, visualizations, interactivity and the integration of models like ChatGPT for enhanced conversational realism. Additionally, in future studies, we want to evaluate the longer-term effects of gamification on user motivation and interaction rates in a real educational context.

Overall, this paper contributes to the growing body of research on gamification in education and lays the foundation for future studies in this field.

Acknowledgements. The research leading to these results has received funding from the German Federal Ministry of Education and Research (BMBF) through the project "Personalisierte Kompetenzentwicklung und hybrides KI-Mentoring" (tech4compKI) (grant no. 16DHB2213).

References

1. Alcasoda, R., Balaoro, J.: Increasing learners' performance and intrinsic motivation using gamified instructional materials: the ARCS model approach. Int. J. Adv. Res. Innovat. Ideas Educ. **8**, 13 (2022)
2. Bezverhny, E., Dadteev, K., Barykin, L., Nemeshaev, S., Klimov, V.: Use of chat bots in learning management systems. In: 4th International Conference on Games and Virtual Worlds for Serious Applications(VS-GAMES 2012), vol. 169, pp. 652–655 (2020)
3. Hodges, C.B.: Designing to motivate: motivational techniques to incorporate in E-learning experiences. Undefined (2004)
4. Colace, F., de Santo, M., Lombardi, M., Pascale, F., Pietrosanto A., Lemma, S.: Chatbot for e-learning: a case of study. Undefined (2018)
5. Davis, F.D.: Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q. **13**(3), 319–340 (1989)
6. Deci, E.L., Ryan, R.M.: Intrinsic motivation and self-determination in human behavior. Springer, Boston (1985). https://doi.org/10.1007/978-1-4899-2271-7
7. Deterding, S., Khaled, R., Nacke, L., Dixon, D.: Gamification: toward a definition. In: CHI 2011 Extended Abstracts on Human Factors in Computing Systems (2011)

8. Fadhil, A., Villafiorita, A.: An adaptive learning with gamification and conversational UIs. In: Bielikova, M., Herder, E., Cena, F., Desmarais, M. (eds.) Adjunct Publication of the 25th Conference on User Modeling, Adaptation and Personalization (UMAP 2017), pp. 408–412. ACM Press, New York (2017)
9. Garrison, D.R.: E-Learning in the 21st Century: A Framework for Research and Practice, 2nd edn. Routledge, New York (2011)
10. Goel, A.K., Polepeddi, L., Watson, J.: A Virtual Teaching Assistant for Online Education (2016)
11. Hoppe, H.U., Joiner, R., Milrad, M., Sharples, M.: Guest editorial: wireless and mobile technologies in education. J. Comput. Assist. Learn. **19**(3), 255–259 (2003)
12. Johnson, C., Urazov, M., Zanoli, E.: Escapeling: a gamified, AI-supported chatbot for collaborative language practice. In: Guralnick, D., Auer, M.E., Poce, A. (eds.) TLIC 2021. LNNS, vol. 349, pp. 141–148. Springer, Cham (2022). https://doi.org/10.1007/978-3-030-90677-1_14
13. Karhulahti, V.M.: Prank, troll, gross and gore: performance issues in esport live-streaming. In: Proceedings of the First International Joint Conference of DiGRA and FDG (DiGRA/FDG 2016). Digital Games Research Association and Society for the Advancement of the Science of Digital Games, Dundee (2016)
14. Keller, J.M.: Development and use of the ARCS model of instructional design. J. Instr. Dev. **10**(3), 2–10 (1987)
15. Marczewski, A.: Even Ninja Monkeys Like to Play: Gamification, Game Thinking and Motivational Design. Gamified, United Kingdom (2015)
16. Neumann, A.T., Conrardy, A.D., Klamma, R.: Supplemental mobile learner support through moodle-independent assessment bots. In: Zhou, W., Mu, Y. (eds.) ICWL 2021. LNCS, vol. 13103, pp. 75–89. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-90785-3_7
17. Park, S.Y.: An analysis of the technology acceptance model in understanding university students' behavioral intention to use e-learning. J. Educ. Technol. Soc. **12**(3), 150–162 (2009)
18. Ryan, D.: Intrinsic and extrinsic motivations: classic definitions and new directions. Contemp. Educ. Psychol. **25**(1), 54–67 (2000)
19. Shi, L., Cristea, A.I.: Motivational gamification strategies rooted in self-determination theory for social adaptive E-learning. In: Micarelli, A., Stamper, J., Panourgia, K. (eds.) ITS 2016. LNCS, vol. 9684, pp. 294–300. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-39583-8_32
20. Siutila, M.: The Gamification of Gaming Streams. GamiFin (2018)
21. de Sousa Borges, S., Durelli, V.H.S., Reis, H.M., Isotani, S.: A systematic mapping on gamification applied to education. In: Cho, Y., Shin, S.Y., Kim, S., Hung, C.C., Hong, J. (eds.) Proceedings of the 29th Annual ACM Symposium on Applied Computing, pp. 216–222. ACM, New York (2014)

A Transfer Learning Approach Interaction in an Academic Consortium

Popescu Doru-Anastasiu[1(✉)], Cristea Daniela-Maria[2], and Bold Nicolae[1,2]

[1] Department of Mathematics and Computer Science, Pitești University Center, National University of Science and Technology POLITEHNICA Bucharest, Bucharest, Romania
dopopan@gmail.com

[2] "1 Decembrie 1918" University of Alba Iulia, Alba-Iulia, Romania
daniela.cristea@uab.ro

Abstract. This paper focuses on presenting a methodology for designing an assessment test that takes into account the degree of difficulty of the items and the overall test structure in an integrated academic environment. This integrated approach allows for the collection of assessment items at an academic consortium level, making them applicable to similar faculties within a group of institutions. The proposed model consists of two main components: the item collector and the test generator. The item collector gathers assessment items from various academic institutions within the consortium. These items are then utilized by the test generator, which automatically generates tests using a combination of evolutionary algorithms (genetic-based) and machine learning (ML) techniques. The genetic-based algorithm and ML methods play a key role in determining the composition and structure of the assessment tests.

Keywords: Assessment consortium · Genetic algorithm · Transfer Learning

1 Introduction

This paper is addressing a research problem about automated test generation and presenting a methodology for designing an assessment test that takes into account the degree of difficulty of the items and the overall test structure in an integrated academic environment. In universities, there are many joint courses, especially in university consortia. The degree of preparation of students can be best determined by tests with the same items. In order to generate assessment tests that can be used in several universities in the consortium, we propose a model web platform containing two components: Assessment - Item Collection (A-IC) and Assessment - Test Generation (A-TG) using ML-based tools. The novelty brought by the paper is related to the usage of an integrated assessment tool within an academic consortium, formed by several academic institutions. This characteristic can help with a standardization of the assessment within the academic environment.

H. Xie et al. (Eds.): ICWL 2023, LNCS 14409, pp. 204–219, 2023.
https://doi.org/10.1007/978-981-99-8385-8_16

Suppose we want to monitor all the assessments inside a consortium and only want to process the questions that relate to courses. There are lots and lots of assessments that are being written each day. We wish to build an automated system (ML classifier) that would filter out all the irrelevant ones, and only forward the appropriate ones to certain courses. In general, when faced with such challenges, we need to either **get training** data or find datasets that we can use for training, and then apply (transfer) the models learned on somewhat different data types. This is sometimes referred to as **transfer learning** although, this phrase also denotes instances where we take a trained model and retrain it on a new kind of data.

The classifier serves a threefold purpose: (1) it employs machine learning techniques to generate keywords for each user-provided question; (2) it leverages machine learning to create question-and-answer groups; and (3) it utilizes genetic algorithms to generate tests.

In this matter, the paper presents several sections regarding the topic. The Literature review Sect. 2 presents state-of-the-art research regarding educational assessment and the usage of genetic algorithms and machine-learning-based tools used for educational purposes. The Model description Sect. 3 presents the description of the model, its components and functionalities, and Results and discussions Sect. 5 related to the implementation of the model for the genetic algorithm test generation with respect to the user requirements. A Conclusions Sect. 6 summarizes the entire research conducted for the paper.

2 Literature

The research in the automated educational assessment area is mainly conducted in several directions, the main two of them being Question Generation (QG) and Answer Evaluation (AE), according to [4]. While the QG research is mainly conducted on the automated generation of questions from a corpus of text using specific language processing methods, the AE research is focused on the analysis of the assessment based on the responses given to the item used for assessment.

Regarding the generation of assessment tests in the context of the existence of several requirements, special attention can be given to the usage of genetic algorithms (GA) with optimal solutions for statistical issues [17], management [5,10] or healthcare [20].

The research assesses competencies needed for university, guiding the development of the RespectNET training program. The matrix of competencies, as outlined in the authors [18] empirical study is categorized into four areas. For each competency, a set of descriptors has been provided, drawing from the European Reference Framework of Competences, the OpenEdu Framework, and the European Framework for the Digital Competence of Educators (DigCompEdu). Additionally, descriptors were collaboratively established by the RespectNET partners consortium. This comprehensive matrix of competence descriptions serves as a foundational resource for the development in designing course modules.

In [9], the author presents a Metaverse education framework in higher education, focusing on a consortium university in Korea. This university has created a shared learning environment through a consortium involving local governments, universities, and companies. They have opened a Metaverse virtual campus to bridge geographical gaps and utilize an LMS system for learning management. Despite differences between the website and Metaverse, the paper outlines suggestions for building a new learning system by leveraging their respective advantages.

In [8], three quality tools were utilized as ad hoc Quality Management (QM) measures for protocol generation/harmonization and data/knowledge exchange, including Sharepoint, Sample Tracker, a self-developed tool for progress monitoring, and Audits, an online auditing process to oversee and assess network activities.

A survey [22] gathered data from around 1,500 students about their academic experiences in fall 2020. It included questions about academic experiences, as well as subsets of questions related to specific topics and modalities of enrollment. The approach of assigning students to answer questions based on their enrollment modality including ensuring relevant feedback, increasing survey completion rates, and allowing variations in questions based on modality types.

The authors [15] discuss a type of optimization algorithm that is inspired by genetics, using notions such as chromosomes, genes, mutation, and crossover. The fitness of a sequence is determined by the number of keywords it shares with the user-defined set of keywords. The chromosomes are ordered by their fitness and the input data is made up of various parameters including the number of tests, the number of generations, and the user-defined keywords. The output data includes the first k solutions and the number of matching keywords for each sequence. In [2], some systems use distributional similarity techniques or extract words from text content for multiple options generation. Other frameworks use deep reinforcement learning for automatic question generation from corpora. The types of generated questions can be categorized by Bloom's taxonomy. Mostow and Chen [13] developed an automated reading tutor that uses automatic question generation to improve students comprehension of text, while only 35.6% of generated questions were deemed acceptable, the accuracy of detecting counterfactual questions was high at 90.0%. Mitkov and colleagues [12] developed a system to generate multiple-choice closed questions with natural languages processing techniques For 1000 question items, the development cost would require 30 h of human work using the system, while 115 h would be required without using the system. In [14], there were used learning materials from a graduate-level introductory data science course at an R1 university in the northeastern United States. The course has six conceptual units and sixteen modules, each consisting of several data science topics. Students also complete seven hands-on coding projects, which are evaluated by an automatic grading system. The authors focused on generating questions from the textual content of the six units in the course using a pipeline. Robinson [19] distinguishes task complexity (cognitive factors), task difficulty (learner factors), and task conditions (interaction factors), meanwhile Campbell [3] contrasted multiple views of complexity, such as a psychological experience, a task-person interaction, and a function of objective characteristics, and created a typology of complex tasks.

In the context of language tasks, Mesmer et al. [11] explicitly distinguish text difficulty (based on the performance of readers) and text complexity (based on textual elements). Beckmann and Goode [1] distinguish apply this distinction to the domain of complex problem-solving, such as controlling a dynamic system with feedback loops.

Gkatzia and Mahamood [6] conducted a study analyzing a corpus of 79 conference and journal papers on NLG published between 2005–2014, which revealed the increasing use of automatic evaluation over human evaluation and intrinsic evaluation over extrinsic evaluation.

3 Model Description

The main variables of the model are the questions, items named after the generation process, the test, the requirements, and the generation mechanism. A question is a particular case of an item, which can consist of a task or an exercise, which is why questions will be considered particular cases of items, and we will refer to questions, tasks, exercises, etc. as items. A set of items with specific properties forms an assessment test (Fig. 1).

The paper focuses on the improvement of the initial model with the automated generation of keywords for each item, which is crucial for the requirement related to the choice of the question based on the subject.

In addition to the three basic components of the model, we also consider the set of constraints (ConSet) formulated by the user, consisting of a list of parameters that can be given various values and based on which these tests are generated. The list of parameters studied so far in previous work includes:

- the subject/subjects that want to be treated in the evaluation;
- the degree of difficulty of the assessment sequence;
- the theoretical/practical report of the evaluation sequence;
- the predominant type of items within the evaluation sequence;
- the maximum time to solve the sequence of items.

Regarding the restriction on the subject of an educational assessment test from the perspective of the subject, there is the problem of determining the main subject of an item, in such a way that it is selected within the assessment test that wants to deal with that subject. In this sense, the determined solution is to describe the item through keywords that best describe that item. Going further with the implementation of the solution, the selection of questions with a certain topic can be done using the user setting the respective topic by a keyword that will filter those items that are described by the chosen keyword. Thus, the problem of describing each item by a series of keywords is raised. This description can be done through the establishment by a human user of these

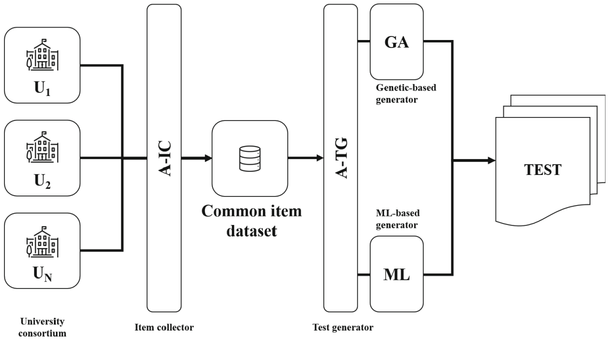


Fig. 1. Visual representation of the model

keywords, but this method is not practical. Therefore, another way of generating these keywords, as well as the process of dynamically extracting some items from a body of text, includes the use of natural language processing (NLP).

To accomplish this, will build a really quick classification model with a simple data pipeline to process the text data, using the Scikit-learn library. Initially, we employ stop-words to remove frequently occurring, non-informative words from our vocabulary. Next, we utilize the term frequency-inverse document frequency (TF-IDF) method to transform the articles into numerical vectors, preserving the importance of terms within the corpus.

The final vectors obtained from the TF-IDF transformation are used to train a Bayesian classifier. The Bayesian classifier is chosen as the first NLP model due to its simplicity and effectiveness in handling text data. The classifier uses probabilistic principles to make predictions based on the features derived from the TF-IDF vectors. To further optimize the performance of the model, we employ grid search to explore different hyperparameter combinations and identify the best configuration for the classifier.

3.1 TF-IDF Digression

The inverse statement frequency is a measure of how much information the word provides, i.e., if it's common or rare across all statements. TF-IDF takes all pieces of text, does a word frequency statistic on them, and then normalizes those frequencies by how common that word is (in general). The TF-IDF (term frequency-inverse document frequency) is defined in [7] as follows:

$$\mathrm{tfidf}(t, d, D) = \mathrm{tf}(t, d) \cdot \mathrm{idf}(t, D) \tag{1}$$

- $\mathrm{tfidf}(t, d, D)$ represents the TF-IDF score of a term t in a statement d within the statement set D.
- $\mathrm{tf}(t, d)$ denotes the term frequency of a term t in a statement d, which measures how frequently the term appears in the statement.

– idf(t, D) stands for the inverse statement frequency of a term t within the statement set D, which quantifies the rarity of the term across the entire set of statements.

In the case of the term frequency TF(t, d), the simplest choice is to use the raw count of a term in a statement, i.e., the number of times that term t occurs in statement d.

3.2 TF-IDF Transformer

In our investigation of the TfidfVectorizer, we performed a TF-IDF transformation on the entire training data, computing the inverse document frequency (IDF) on all statements, which resulted in an internal state of IDF values. Each statement is represented as a vector in a high-dimensional space, where each feature corresponds to a unique word in the corpus.

Later, when we applied the tf.transform() method to an individual statement, it computed the term frequencies (TF) for that specific statement and multiplied them with the precomputed IDF values from the internal state. This yielded the TF-IDF scores for all words in the vocabulary, corresponding to the given statement.

When we examined the result of this transformation, we encountered a sparse matrix representation (1, 3113) for a single statement, signifying 3113 TF-IDF features or values, one for each word in the vocabulary. As expected, the vast majority of these values were 0, as each statement typically contains only a few of the possible words.

4 Algorithms

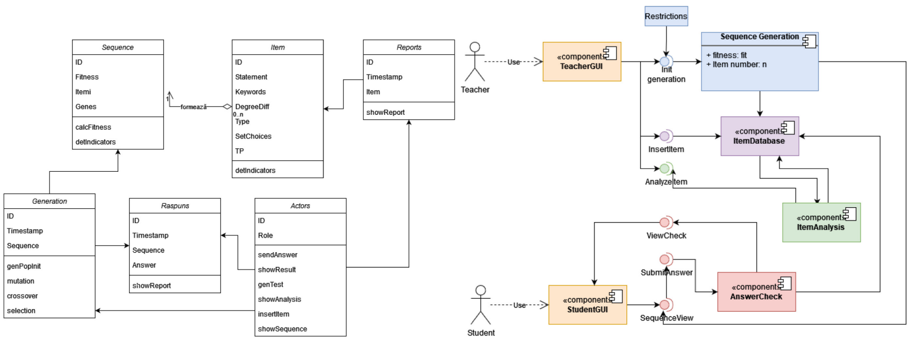
4.1 Item Collection (A-IC)

A-IC Components. The item collection (A-IC) component is the one responsible for the creation of a dataset of items and the continuous improvement and update of the items present in the dataset. The model is based on the existence of the item and its characteristics. In this matter, the item, which will be denoted further by q (id_q; st_q; dd_q; v_q; t_q, ans_q) is an object formed of the next components:

- the identification part of the item id_q, which consists of an identification particle of the item; in our model, id_q is considered to be a number and will play an important role in the implementation phase ($id_q \geq 1$, $id_q \in$ N);
- the statement st_q, formed of the phrase or a set of phrases that contain the requests of the item (either questions, statements, or exercises); the statement is an important part regarding the automated generation of tests;
- the set of keywords kw_q, of cardinal nr_{kw}, which contains the list of keywords that describe the item, either its category or the main topic of the item. A keyword kw_{q_i}, $i = 1$, nr_{kw}, can be established by a human operator, or it can be generated using specific methods of language programming, such as Natural Language Programming (NLP);

- the degree of difficulty dd_q, $dd_q \in [0, 1]$; which is firstly pre-established using the method presented in [16]. The degree of difficulty is then calculated as the ratio between the number of incorrect answers at the item and the total number of answers, being continuously updated by the answers given to the item;
- the item type t_q, $t_q \in \{\text{'m','s','e'}\}$, which shows the type of the item, whether it is multiple-choice based (m) or the answer is a textual one, given by the user, in case of short (s) and essay (e) types.
- choices set v_q (when $t_q =$"m"), which can be formed of a list of two or more possible answers when the item type is multiple or is null when the item type is not ($v_q \geq 1$, $v_q \in \mathrm{N}$);
- the correct answer of the item ans_q, which contains the correct answer of the item; it has the form of:
  - a choice identifier, such as a letter (a, b, c, ...) or a number (1, 2, 3, ...), in case of multiple-choice items ($t_q =$ "m");
  - a real number, in case of numerical answers;
  - a text, in case of short or essay items ($t_q \in \{\text{'s', 'e'}\}$).

The component diagram, shown in Fig. 2b, emphasizes the modular character of the program system, starting from the three main components of the general model. The diagram presents aspects related to restrictions, interfaces, databases, and actors involved in the system (Teacher and Student). The component diagram helps to visualize the behavior of the system in the context of a real-world application. Also, the Item class shows the modality of the storage of the generated items in the database, as shown in Fig. 2a.



(a) The Item class and its relation with the other classes

(b) Arhitecture model, comprising the item generation and storage

Fig. 2. Model arhitecture

4.2 Test Generation (A-TG)

A-TG Components. An important component of the test generation model is a test T (S, DD), which is considered a set of items q_i, $i = 1, \|S\|$, where S is the set of items that form the test, dd_i is the degree of difficulty of the item i and DD is the degree of difficulty of the test:

$$DD = \sum_{i=1}^{S} q_{dd_i} \tag{2}$$

A-TG Functionality

A-TG Using Genetic Algorithms. The pseudocode of the genetic algorithm (GA) is presented in Algorithm 1.

Algorithm 1 Genetic Algorithm for A-TG

Input: $population\_size, generations$
Output: $\mathbf{X^*}(itemIDsequence)$
Population initialisation $X_i (i = 1, 2...n)$
Calculate the fitness value of every chromosome
Selection
$\mathbf{X^*} \longleftarrow$ the best chromosome
$t \longleftarrow 0$
while $(t < generations)$ **do**
  for (every chromosome) **do**
    $\mathbf{X}\,(t+1) \longleftarrow Mutation(\mathbf{X}(t), \mathbf{Y}(t), \mathbf{X^*})$
    $\mathbf{X}\,(t+1) \longleftarrow Crossover(\mathbf{X}(t), \mathbf{X^*})$
  end for
  Calculate the fitness of the modified chromosome;
  Selection
  $\mathbf{X^*} \longleftarrow$ the best chromosome with the minimum fitness
  $t \longleftarrow t + 1$
end while
return $\mathbf{X^*}$

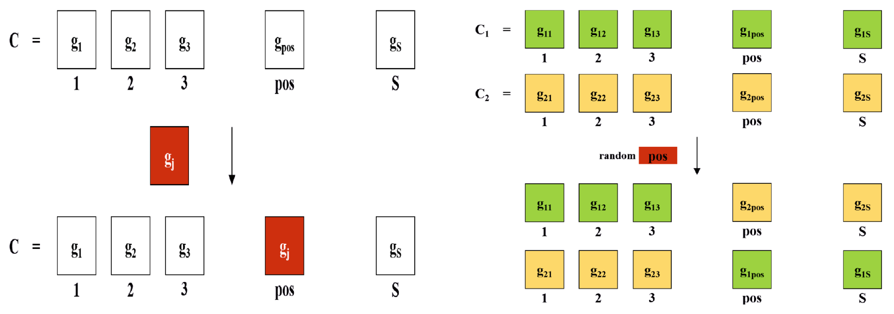
The genetic algorithm codifies items as genes within a chromosome, which will represent a test. In this matter, the next components will be needed:

- a chromosome C, which codifies a test, formed of the next components:
  - order number id, $id \in \{0, \ldots, NrI\}$, where NrI is the total number of items within the database;
  - the gene set $G_j = \{g_j | j \in \{1, \ldots, S\}\}$, where $G = S$; $j = 1, nr_T$, $nr_T = n_1 + n_2 + ... + n_i$, $i = 1$, n;
  - the fitness function f, which calculates the closeness of the degree of difficulty of the chromosome C to the desired degree of difficulty DD. The fitness function calculation also gives the flexibility of choosing a lower or a higher level of difficulty, a higher value of min meaning a higher level of difficulty, and a lower value of min combined with a lower

value of *max* meaning a lower level of difficulty. The fitness function is defined as follows:

$$f(C) = \left| \sum_{i=1}^{S} q_{dd_i} - \frac{min + max}{2} \right| \quad (3)$$

– the genetic operators, the ones used in the presented model being:
  - the generation of a chromosome $GenC$, an operator that forms a new chromosome;
  - the mutation of a chromosome $MutC$, which can be defined as follows: for a given chromosome C, a random position within the chromosome pos, $pos = 1, S$, and a randomly selected gene g_j, $j = 1, NrI$, the mutation operation is defined as the shift of the gene g_{pos} with the gene g_j.
  - the crossover of a chromosome $CsvC$, which can be defined as follows: for two given chromosomes C_1 and C_2 and a randomly selected position within the two chromosomes pos, the first part of the chromosome C_1 up to the gene g_{pos} is combined with the second part of the chromosome C_2 from the gene g_{pos} to the end of C_2 and the first part of the chromosome C_2 up to the gene g_{pos} is combined with the second part of the chromosome C_1 from the gene g_{pos} to the end of C_1, resulting in two new chromosomes C'_1 and C'_2.



(a) The $MutC$ operator (b) The $CsvC$ operator

Fig. 3. Visual representation of the operators

A-TG Using ML-Based Tools - Measurement Framework Architecture. In order to compare the metrics of multiple models consistently, we have established a measurement framework to allow us to assess their performance. The core component is a measure (model, x, y) function that takes the model, accepts a list of strings as input (x), and returns a list of numbers representing the predictions with the correct labels (y) adhering to a classification report regarding the metrics.

Model 1. As part of the implementation process, we have created our own model 1 function that serves as a wrapper for the pipeline model. The prediction function takes a list of input texts (x), tokenizes it into words, checks if any of the words intersect with the predefined parameter sets and returns a list of predictions associated with these texts. In our case, the predictions are the classifications resulting from the classifier model on the test dataset (classification_test_inputs).

Model 2. Nevertheless, some optimisations are required to find the best overall model based on the prediction classifications resulting from the GridSearchCV model.

Model 3. The previous model used all the features (words) in the vocabulary, but realistically when dealing with lots and lots of files, the vocabulary will increase to hundreds of thousands of words and thus each TF-IDF entry will be a vector of size (1, x00 000), ridiculously large. This is sometimes called "the dimensionality curse" and it happens when on small scale things seem to be working but in a production environment, the size of the data starts to make the approach intractable. In order to solve this, we will be using an improvement over the previous model, namely we will not keep the full vocabulary while computing the results, but only some of the most relevant words getting the SelectKBest instance from the pipeline where all the words that the feature selection step has chosen and printed all the selected words, in descending order of their score (the most important ones are at the top).

Model 4. If we are looking closely, there are many words that have the same meaning but are slightly different (because of the context in which they are used). In an example resulting from the dataset, we have "thanks" and "thanksgiving", so we would like to normalize them as much as possible to keep the overall vocabulary small. This process can be accomplished through two approaches, *stemming* and *lemmatization*. So far, we implemented our own transformer and do both at the same time. NLTK already provides us with convenient objects that we can use, where we can inject this class into the TfidfVectorizer as a custom tokenizer function and end up a mixed object where we won't be able to use the stop_words = 'english' parameter because it will be applied to that list as well. Considering this, we got a standalone StopWordsTransformer that will filter out the stop-words and a standalone LemmaStemmerTransformer that should take the non-stop words and do lemmatization and semisation on them. These two custom transformers should be placed in front of the TfidfVectorizer.

The initial model performances yielded F1-scores of **0.79** for the first model and **0.74** for the second one. Upon invoking the measure function with the test data (classification_test_inputs) and correct labels (classification_test_labels), all three models achieved perfect classification for each class, resulting in a harmonic mean **F1-score** of **1**. The dataset is well-balanced, as it contains a similar number of instances for each class (Fig. 3).

5 Results and Discussions

Firstly, we will showcase results for items with manually set keywords, as presented in Table 1. Secondly, we will present results for items with automatically set keywords, as shown in Table 2. The shown tables present similar results, with slightly finer results for the automated generated keywords, given by the user refinement. In order to show the effectiveness of the NLP automated extraction of the keywords, we have used the WordNet library to automatically extract two keywords from each statement of the questions. The compared versions of the keywords are shown in Table 3.

A great deal of importance must be brought to refining the generation of keywords, with the additional feature of generating synonyms to the found keywords. Also, it can be observed that there are significant differences between the user keywords and the generated keywords, due to the close relationship between the vocabulary of the statement and the generated keywords. This aspect can be improved by refining the algorithm with the addition of finding synonyms or words close to the meaning of the ones in the statement.

Essentially for optimizing the model accuracy, we explored some techniques to work effectively with numerical data and algorithms to enhance the performance of the model as presented in Table 4, which has been transformed into a numerical vector format.

Feature Selection. Based on the experiments, we employed the SelectKBest and SelectPercentile methods to identify the most informative features from our dataset, where was achieved an accuracy of **0.53**.

Ensemble Methods. To further enhance the accuracy, we leveraged the power of ensemble methods such as Random Forest and Gradient Boosting, where was obtained an improved accuracy of **0.6**.

Hyperparameter Tuning. By using GridSearchCV, we fine-tuned the hyperparameters of our model to identify the optimal combination, resulting in an accuracy of **0.67** and demonstrating the importance of selecting appropriate hyperparameters for optimal model behavior.

Bayesian Classifier. We implemented a Bayesian classifier on top of the final vectors obtained from the TfidfTransformer and SelectKBest. The MultinomialNB classifier was utilized and got a comprehensive classification report with an overall score of **0.57**.

K-Means Clustering. For a deeper understanding of the data, we applied K-Means clustering on the TF-IDF matrix of the text data. This technique allowed us to assign data points to different clusters, where the best estimator achieved a score of **0.67**, indicating the effectiveness of K-Means clustering in grouping similar data points together.

Figure 4a shows that the best values for the closeness of the degree of difficulty can be found for a larger initial population size. In this matter, minimum and maximum values are the lowest for an initial population size of 50 from the ones taken into consideration.

Figure 4b shows that the best values for the closeness of the degree of difficulty can be found for a lower number of generations. In this matter, minimum and maximum values are the lowest for a number of 200 generations from the ones taken into consideration.

Table 1. Obtained values of the test for the input data for the manual setting of the keywords

ID	Statement	University	Diff
57	What is the result of the expression: $(100 = 101)$ AND $(2*3 < 6)$	UPIT	0.54
65	Which function in Excel returns the sum of a range of numbers?	UPIT	0.34
21	What given number is prime?	UPIT	0.34
8	Which of these is not an item of hardware to do with computers?	UAI	0.99
42	What is wrong with the following piece of code: includ ?	UAI	0.89
55	What is the result of the expression: $(5 > 7)$ OR $(0 < 2 * 5 < 15)$	UPIT	0.87
45	What is the name of the direction of a page used for viewing and printing?	UPIT	0.17
17	What is a URL?	UPIT	0.38
13	Which is the best application to use to write a letter?	UPIT	0.16

Table 2. Obtained values of the test for the input data

ID	Statement	University	Diff
6	The Microsoft Word program is	UAI	0.34
10	What is the name of the direction of a page used for	UPIT	0.35
7	The extension of a file created in Word is:	UAI	0.82
9	The process of removing an unwanted part of an image	UAI	0.99
12	What is the term for unsolicited emails?	UPIT	0.59
66	What is a TCP/IP?	UPIT	0.95
14	URL means:	UPIT	0.16
23	The process of arranging the elements of a column...:	UPIT	0.65
23	Which function in Excel returns the average...	UPIT...	0.16

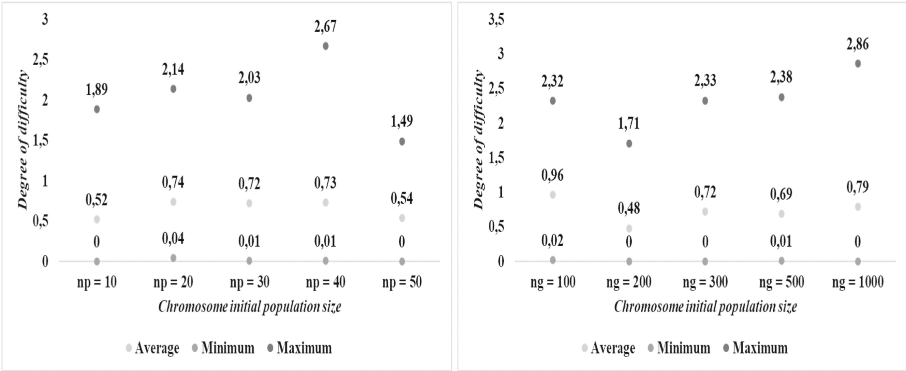
Table 3. Example of automated generated keywords for random 10 questions in the database

Item ID	User keywords	Generated keywords
1	operating, system	operating, following, system
2	operating, system	operating, mobile, system
3	operating, Windows	systems, differences, windows, linux
4	path, file	path, desktop, file, hello.txt
5	operating, Linux	operating, words, system, linux
6	C++, library	library, files
7	string, palindrome	code, string, palindrome
8	Android, code	code, wrong, android
9	encryption, data	data, confidential
10	hardware, item	item, hardware, computers

Table 4. Model Performance Experiments

Experiment	Accuracy	Notes
Feature Selection	0.53	Statistical relationship between features and target variable considered
Ensemble Methods	0.6	Random Forest and Gradient Boosting used for model combination
Hyperparameter Tuning	0.67	GridSearchCV and RandomizedSearchCV to find optimal hyperparameters
Bayesian Classifier	0.57	MultinomialNB classifier on TfidfTransformer and SelectKBest vectors
K-Means Clustering	0.67	K-Means applied on TF-IDF matrix of text data

In real-world scenarios, datasets [21] often exhibit class imbalance, even if are well-balanced, with a similar number of instances for each class. In this case, the F1-score metric does not accurately reflect the model performance. The inefficiency of traditional models in handling enormous TF-IDF vectors is emphasized and results demonstrate the importance of careful feature selection, ensemble methods, hyperparameter tuning, and clustering techniques to achieve higher accuracy and performance in our model. These findings are crucial for enhancing the overall quality and reliability of our model, making it more suitable for real-world applications.



(a) The variation based on the initial population size

(b) The variation based on the number of generation NG

Fig. 4. The variation of the degree of difficulty

6 Conclusions

In this work we conducted an experiments to evaluate a method for assessment test generation in an academic consortium. The proposed model consists of two components. The first component, namely, A-IC, aims to collect items for the assessment and creating the item database. The second component, namely, A-TG, aims at generating the test using genetic algorithms and establish a measurement framework to compare multiple models. In particular, Bayesian classifier is employed to classify the item based on the TF-IDF representation.

A great deal of importance must be brought to refining the generation of keywords, with the additional feature of generating synonyms to the found keywords. Also, it can be observed that there are significant differences between the user keywords and the generated keywords, due to the close relationship between the vocabulary of the statement and the generated keywords. This aspect can be improved by refining the algorithm with the addition of finding synonyms or words close to the meaning of the ones in the statement.

Acknowledgment. Cristea Daniela Maria acknowledges the support from the Romanian Ministry of Education, under contract no. 3721/2023, and of the University "1 Decembrie 1918" of Alba Iulia https://en.uab.ro/centre/1-center-for-scientific-research/

References

1. Beckmann, J., Birney, D., Goode, N.: Beyond psychometrics: the difference between difficult problem solving and complex problem solving. Front. Psychol. **8**, 1739 (2017). https://doi.org/10.3389/fpsyg.2017.01739
2. Blšták, M., Rozinajova, V.: Automatic question generation based on sentence structure analysis using machine learning approach. Nat. Lang. Eng. **28**, 1–31 (2021). https://doi.org/10.1017/S1351324921000139
3. Campbell, D.: Task complexity: a review and analysis. Acad. Manag. Rev. **13**, 40–52 (1988). https://doi.org/10.5465/AMR.1988.4306775
4. Das, B., Majumder, M., Phadikar, S., Sk, A.A.: Automatic question generation and answer assessment: a survey. Res. Pract. Technol. Enhanc. Learn. **16**, 1–5 (2021). https://doi.org/10.1186/s41039-021-00151-1
5. Davis, L.: Job shop scheduling with genetic algorithms. In: Proceedings of the First International Conference on Genetic Algorithms and their Applications, pp. 136–140. Psychology Press (2014)
6. Gkatzia, D., Mahamood, S.: A snapshot of NLG evaluation practices 2005–2014 (2015). https://doi.org/10.18653/v1/W15-4708
7. Hamdaoui, Y.: TF(term frequency)-IDF(inverse document frequency) from scratch in python. Towards Data Sci. **23**, 2022 (2019). https://towardsdatascience.com
8. Hülsemann, M., et al.: Introducing quality measures in an academic research consortium: lessons and recommendation from implementing an ad hoc quality management system for organ model research. EMBO Rep. **23**, e55095 (2022). https://doi.org/10.15252/embr.202255095
9. Jeong, Y., Choi, S., Ryu, J.: Work-in-progress-design of LMS for the shared campus in metaverse learning environment. In: 2022 8th International Conference of the Immersive Learning Research Network (iLRN), pp. 1–3 (2022). https://doi.org/10.23919/iLRN55037.2022.9815909
10. Lee, C.K.H.: A review of applications of genetic algorithms in operations management. Eng. Appl. Artif. Intell. **76**, 1–12 (2018)
11. Mesmer, H., Cunningham, J., Hiebert, E.: Toward a theoretical model of text complexity for the early grades: learning from the past, anticipating the future. Read. Res. Q. **47**, 235–258 (2012). https://doi.org/10.1002/rrq.019
12. Mitkov, R., Ha, L., VARGA, A., Rello, L.: Semantic similarity of distractors in multiple-choice tests (2009). https://doi.org/10.3115/1705415.1705422
13. Mostow, J., Chen, W.: Generating instruction automatically for the reading strategy of self-questioning (01 2009). https://doi.org/10.3233/978-1-60750-028-5-465
14. Nguyen, H.A., Bhat, S., Moore, S., Bier, N., Stamper, J.: Towards generalized methods for automatic question generation in educational domains (2022). https://doi.org/10.1007/978-3-031-16290-920
15. Popescu, A.D., Bold, N., Nijloveanu, D.: A method based on genetic algorithms for generating assessment tests used for learning. Polibits **54**, 53–60 (2016)
16. Popescu, D.A., Bold, N.: The development of a web application for assessment by tests generated using genetic-based algorithms. In: CEUR Workshop Proceedings (2016)
17. Quirós, P., Lasheras, F.S.: Methodology for the projection of population pyramids based on monte carlo simulation and genetic algorithms. Appl. Intelligence pp. 1–18 (2023)
18. Robescu, D., Reiner, S., Trunk, A.: Toward a matrix of competences for respectful communication in the university-civil society context (2023). https://doi.org/10.21125/edulearn.2023.0217

19. Robinson, P.: Task complexity task difficulty and task production: exploring interactions in a componential framework. Appl. Linguist. **22**, 27–57 (2001). https://doi.org/10.1093/applin/22.1.27
20. Sharma, S., Kumar, V.: Application of genetic algorithms in healthcare: a review. Next Generation Healthcare Informatics, pp. 75–86 (2022)
21. Tiwari, R.: Imbalanced learning in fraud prevention: Views and solutions from the trenches. CueNex, February 2023. https://medium.com/cuenex, February 2, 2023
22. Whelehan, D.: Students as partners: a model to promote student engagement in post-COVID-19 teaching and learning. J. Higher Educ. (2020)

Author Index

H. Xie et al. (Eds.): ICWL 2023, LNCS 14409, pp. 221–222, 2023.
https://doi.org/10.1007/978-981-99-8385-8