Cognition and Exploratory Learning in the Digital Age

Dirk Ifenthaler
Pedro Isaías
Demetrios G. Sampson *Editors*

Orchestration of Learning Environments in the Digital World

Springer

Cognition and Exploratory Learning in the Digital Age

This book series focuses on the implications of digital technologies for educators and educational decision makers. Taking an interdisciplinary and visionary perspective on cognition and exploratory learning in the digital age, the series examines how digital technologies can or should be exploited to address the needs and propel the benefits of large-scale teaching, learning and assessment. It includes the newest technical advances and applications, and explicitly addresses digital technologies and educational innovations.

The series incorporates volumes stemming from the Cognition and Exploratory Learning in the Digital Age (CELDA) conference. Since 2004 the CELDA conference has focused on discussing and addressing the challenges pertaining to the evolution of the learning process, the role of pedagogical approaches and the progress of technological innovation, in the context of the digital age. Some of CELDA's main topics include: assessment of exploratory learning approaches and technologies, educational psychology, learning paradigms in academia and the corporate sector, student-centred learning and lifelong learning. In line with and in addition to these CELDA conference volumes, this series publishes monographs and edited volumes covering the following subject areas:

- K-12, Higher Education, Vocational Education, Business Education, Health Education
- Learning Design, Instructional Design
- Change Management, Technology Adoption, Learning Organization
- Data analytics
- Learning, Teaching, Assessment

The book series represents the most recent work in the area where cognitive psychology, information and communications technology and instructional design intersect.

More information about this series at http://link.springer.com/series/16424

Dirk Ifenthaler • Pedro Isaías
Demetrios G. Sampson
Editors

Orchestration of Learning Environments in the Digital World

Springer

Editors
Dirk Ifenthaler
Learning, Design and Technology
University of Mannheim
Mannheim, Germany

UNESCO Deputy Chair on Data Analytics
in Higher Education Learning and Teaching
Curtin University
Perth, WA, Australia

Demetrios G. Sampson
Department of Digital Systems
University of Piraeus
Athens, Greece

Pedro Isaías
Information Systems & Technology
Management School (ISTM)
The University of New South Wales
(UNSW – Sydney)
Sydney, NSW, Australia

ISSN 2662-5628 ISSN 2662-5636 (electronic)
Cognition and Exploratory Learning in the Digital Age
ISBN 978-3-030-90943-7 ISBN 978-3-030-90944-4 (eBook)
https://doi.org/10.1007/978-3-030-90944-4

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

The Cognition and Exploratory Learning in the Digital Age (CELDA; www.celda-conf.org) conference focuses on discussing and addressing the challenges pertaining to the evolution of learning processes, the role of pedagogical approaches, and the progress of technological innovation, in the context of the digital age. In each edition since 2004, CELDA gathers researchers and practitioners in an effort to cover both technological and pedagogical issues in groundbreaking studies. Some of CELDA's main topics include: assessment of exploratory learning approaches and technologies, educational psychology, learning paradigms in academia and the corporate sector, student-centered learning and lifelong learning. The 2020 edition of the CELDA conference received a wide array of papers examining the deployment of learning technologies, proposing pedagogical approaches and practices to address digital transformation, presenting case studies of specific technologies and contexts, and overall debating the contribution of learning technologies for the improvement of the learning process and the experience of the students and for the development of key competences.

A common theme of the CELDA 2020 conference was the design of learning environments. Yet, the design of learning environments is not a simple task. Rather, it includes a systematic analysis, planning, development, implementation, and evaluation phases (Ifenthaler, 2012). However, the increased availability of digital media and technology for learning and teaching requires a continuous orchestration of learning environments in the digital world. This edited volume *Orchestration of Learning Environments in the Digital World* provides a platform for the continuous conversation stemming from the CELDA 2020 conference series. It comprises two parts focusing on different stakeholder groups and cases: Part I—Learning Environments Focused on Schools; Part II—Learning Environments Focused on Higher Education and Further Education.

In Part I, the first chapter "Adaptive Exercises and Formative Assessment for English Remedial Action" concerns the design and implementation of activities and resources within digital learning environments, their use by students, and their effectiveness in improving language skills (Elisa Corino, Cecilia Fissore, Marina Marchisio, Chap. 1). The next chapter "Learning How to Use a Digital Workbench:

Guided or Explorative?" investigates the influence of the two instructional approaches on different learning objectives (Josef Buchner, Anne-Cathrin Vonarx, Peter Pfänder, Michael Kerres, Chap. 2). The concluding chapter in Part I, "Longitudinal Co-teaching Projects: Scoping Review," aims to provide an overview of approaches not only to the actual implementation of co-teaching, but also to the whole process of preparation in order to evaluate the experience gained in a number of researches already carried out, in the form of an overview study (Jaroslav Veteska, Martin Kursch, Zuzana Svobodova, Michaela Tureckiova, Lucie Paulovcakova, Chap. 3).

In Part II, the first chapter "Characterizing Personal Educational Goals" focusses on individual educational goals, which are central to self-regulated learning and an observable manifestation of students' motivation (Felix Weber, Tobias Thelen, Chap. 4). Next, "A Digital Environment for University Guidance" aims at facilitating the transition from secondary schools to higher education with an open platform that delivers automatic assessments (Francesco Floris, Marina Marchisio, Sergio Rabellino, Matteo Sacchet, Chap. 5). The following chapter "Examining the Influence of Cognitive Ability on Situating to a Video Game" examines the relationship between an individual player's cognitive ability and their in-game experience (Michael P. McCreery, Danielle L. Head, Sam A. Leif, Joseph P. Fiorentini, Catherine A. Bacos, Jeffrey R. Laferriere, S. Kathleen Krach, Le Quanda L. Cole, Chap. 6). Then, "Study Properly: Digital Support for the Pre-student Phase" presents a project to support the choice of study direction (Pia Drews, Alke Martens, Chap. 7). Next, "Investigating How Word Clutter and Colour Impact Upon Learning" presents a recognition experiment whereby participants had to learn target words that were presented in black or red font and positioned among no clutter, clutter words (distractor words surrounded the target), and clutter non-words (sequence of random letters surrounded the target) (Olivia Foulds, Chap. 8). Then, "Evaluating the Spatial Continuity Effects of Augmented Reality System on Learning Performance and Psychological Factors" performed comparative analyses on two mobile augmented reality learning systems based on the learning performance, cognitive load, and technology acceptance using the spatial continuity principle (Xuewang Geng, Masanori Yamada, Chap. 9). In their chapter "Social Robots in Education," Josef Guggemos, Sabine Seufert, Stefan Sonderegger, and Michael Burkhard aim to provide a conceptual overview of social robots for education (Chap. 10). Next, "Undergraduate Mathematics Students Engaging in Problem-Solving Through Computational Thinking and Programming" aims at exploring students' engagement in mathematical problem-solving through computational thinking (CT) and the programming language MATLAB (Said Hadjerrouit, Nils-Kristian Hansen, Chap. 11). The following chapter "Summary Writing with Graphic Organizers in Web-Based Investigative Learning" presents a cognitive tool that scaffolds the learning process according to the learning model (Oriko Harada, Akihiro Kashihara, Chap. 12). Then, "Interaction Preferences in Digital Learning Environments: Does Gender and Achievement Matter?" focuses on the interaction preferences of bachelor students in a digital learning environment based on their gender and achievement situation (Muhittin Sahin, Dirk Ifenthaler, Chap. 13).

Further, “Development of Fill-In Workbook System to Visualize Learning Attitude” (Kousuke Abe, Tetsuo Tanaka, Kazunori Matsumoto, Chap. 14) presents a system which provides an authoring mode in which teachers create teaching materials, a presentation mode in which teachers explain teaching materials while referring to the students’ learning attitude, a browsing mode in which students refer to teaching materials while filling in the blanks, and an analytics mode to support a retrospective review of teachers and students. The following chapter “MAI Helper - Learning Support System for Time Management Skill Acquisition using Learning Analytics” presents a system based on a learning analytics approach with functions that allow students to make their learning schedules and reflect on their learning behaviors (Hiroyuki Watanabe, Li Chen, Xuewang Geng, Yoshiko Goda, Atsushi Shimada, Masanori Yamada, Chap. 15). The concluding chapter in Part II, “The Development and Validation of a Concept Mapping Emotions Scale (Cm-ES) for University Students,” aims to develop and verify a concept mapping emotions scale and to investigate university students’ emotions of concept mapping with this scale (Xinyuan Li, Yanyan Duan, Tong Lu, Guoqing Zhao, Chap. 16).

The contributions and outcomes collected in this edited volume are consistent with the overall goal of the Cognition and Exploratory Learning in the Digital Age book series (www.springer.com/series/16424) and further document the advances presented and published from previous editions of the CELDA conference. In their first publication, Spector et al. (2010) approach the general developments and challenges of learning and instruction in the digital age. More specifically, the editors gathered contributions that examined cognitive approaches to learning and instruction, knowledge representation and mental models technology, facilitated tools and techniques, communications and methods, and integrative methods and online learning. In Ifenthaler et al. (2011), the editors compiled research initiatives that emphasize multiple perspectives on problem solving and learning in the context of the digital age by exploring related topics such as pedagogical usability issues in web-based learning objects, automated measurement of critical thinking for discussion forum participants, expanding global awareness with virtual collaboration, and simulation games as learning experience. In Isaias et al. (2012), the editors intended to assess the impact of web 3.0 in learning and instruction by focusing on student-centered learning, collaborative learning, and exploratory technologies, and addressing educational precepts such as just-in-time learning, constructivism, and web 3.0’s adoption in education. Following the tendency for the adoption of mobile devices in education, Sampson et al. (2013) compiled the most relevant contributions pertaining to ubiquitous and mobile learning in the digital age and all its fundamental ramifications, such as formal and informal learning environments, social web technologies, virtual worlds and game-based learning, and location-based and context-aware environments. On a later publication Sampson et al. (2014) emphasized the importance of digital systems for open access in the context of both formal and informal learning and gathered contributions that covered the theoretical and practical aspects of open access, as well as different methods and technologies used to support it. In Isaias et al. (2015) the focus was placed on e-learning systems, which were scrutinized from different perspectives: exploratory learning technologies, e-learning social web design, learner communities through e-learning implementations, and collaborative and student-centered e-learning design. In the following year, Spector et al. (2016) gathered contributions about the competencies, challenges, and transformation that stem from the deployment of digital technologies. The publication introduces this subject, reflects about the changes in learning and instructional paradigms, debates assessments and analytics for teachers and decision makers, and examines the changing tools and environments teachers and learners must face. In Sampson et al. (2018), digital technologies were explored from the perspective of their role as promoters of sustainable educational innovations for the enhancement of teaching, learning, and assessment in all educational levels. The research depicted in this publication addressed the importance of digital technologies in transforming the learning environment, enriching the student learning experiences, measuring and assessing teaching and learning, and cultivating student competences for the digital smart society. Then, Sampson et al. (2019) focused on the transformational potential that learning technologies have for large-scale teaching, learning, and assessment. The editors gathered the outcomes of research efforts featuring state-of-the-art case studies examining the innovative influence of learning technologies, such as Massive Open Online Courses and educational data analytics. Two additional volumes have been published exploring the conceptual and practical aspects of technologies that are used to support learning, with a multidisciplinary approach that encompasses all levels of education (Isaias et al., 2020a, b). Ifenthaler et al. (2021) addressed the tension between digital technologies and learning sciences. The contributions of the volume highlight the need for a critical discourse in educational technology linked to a large body of knowledge from learning sciences.

The CELDA conferences (www.celda-conf.org) and related book series (www.springer.com/series/16424) strive to continue the critical discourse regarding the support of learning processes and learning outcomes through digital technologies. This edited volume adds to these continuous conversations and documents the advances of our field.

Mannheim, BW, Germany
Perth, WA, Australia — Dirk Ifenthaler
Sydney, NSW, Australia — Pedro Isaías
Athens, Greece — Demetrios G. Sampson

References

Ifenthaler, D. (2012). Design of learning environments. In N. M. Seel (Ed.), *Encyclopedia of the sciences of learning* (Vol. 4, pp. 929–931). Springer. https://doi.org/10.1007/978-1-4419-1428-6_186.

Ifenthaler, D., Kinshuk, Isaias, P., Sampson, D. G., & Spector, J. M. (Eds.). (2011). *Multiple perspectives on problem solving and learning in the digital age*. Springer.

Ifenthaler, D., Sampson, D. G., & Isaias, P. (Eds.). (2021). *Balancing the tension between digital technologies and learning sciences*. Springer.
Isaias, P., Ifenthaler, D., Kinshuk, Sampson, D. G., & Spector, J. M. (Eds.). (2012). *Towards learning and instruction in Web 3.0. Advances in cognitive and educational psychology*. Springer.
Isaias, P., Sampson, D. G., & Ifenthaler, D. (Eds.). (2020a). *Online teaching and learning in higher education*. Springer.
Isaias, P., Sampson, D. G., & Ifenthaler, D. (Eds.). (2020b). *Technology supported innovations in school education*. Springer.
Isaias, P., Spector, J. M., Ifenthaler, D., & Sampson, D. G. (Eds.). (2015). *E-Learning systems, environments and approaches*. Springer.
Sampson, D. G., Ifenthaler, D., Spector, J. M., & Isaias, P. (Eds.). (2014). *Digital systems for open access to formal and informal learning*. Springer.
Sampson, D. G., Ifenthaler, D., Spector, J. M., & Isaias, P. (Eds.). (2018). *Digital technologies: Sustainable innovations for improving teaching and learning*. Springer.
Sampson, D. G., Isaias, P., Ifenthaler, D., & Spector, J. M. (Eds.). (2013). *Ubiquitous and mobile learning in the digital age*. Springer. https://doi.org/10.1007/978-1-4614-3329-3.
Sampson, D. G., Spector, J. M., Ifenthaler, D., Isaias, P., & Sergis, S. (Eds.). (2019). *Learning technologies for transforming teaching, learning and assessment at large scale*. Springer.
Spector, J. M., Ifenthaler, D., Isaias, P., Kinshuk, & Sampson, D. G. (Eds.). (2010). *Learning and instruction in the digital age*. Springer.
Spector, J. M., Ifenthaler, D., Sampson, D. G., & Isaias, P. (Eds.). (2016). *Competencies in teaching, learning and educational leadership in the digital age*. Springer.

Acknowledgments

We acknowledge the support of the International Association for Development of the Information Society (http://www.iadisportal.org/) and for its continuing sponsorship of the ***International Conference on Cognition and Exploratory Learning in Digital Age*** (CELDA; http://www.celda-conf.org). In addition, we wish to express our gratitude to all the members of the CELDA 2020 international program committees for their timely and thoughtful reviews of all papers submitted to the conference. We also would like to acknowledge Springer Science + Business Media and Springer's education editor, Melissa James, for supporting the publication of this edited volume based on expanded and edited versions of the best CELDA 2020 papers.

Mannheim, BW, Germany
Perth, WA, Australia
Sydney, NSW, Australia
Athens, Greece

Dirk Ifenthaler
Pedro Isaías
Demetrios G. Sampson

Contents

The original version of this book was revised. The correction is available at https://doi.org/10.1007/978-3-030-90944-4_17

About the Editors

Dirk Ifenthaler is Professor and Chair of Learning, Design and Technology at the University of Mannheim, Germany, and UNESCO Deputy Chair of Data Science in Higher Education Learning and Teaching at Curtin University, Australia. His previous roles include Professor and Director, Centre for Research in Digital Learning at Deakin University, Australia, Manager of Applied Research and Learning Analytics at Open Universities, Australia, and Professor for Applied Teaching and Learning Research at the University of Potsdam, Germany. He was a 2012 Fulbright Scholar-in-Residence at the Jeannine Rainbolt College of Education, at the University of Oklahoma, USA. Dirk's research focuses on the intersection of cognitive psychology, educational technology, data analytics, and organizational learning. His research outcomes include numerous coauthored books, book series, book chapters, journal articles, and international conference papers, as well as successful grant funding in Australia, Germany, and USA. He is the Editor-in-Chief of *Technology, Knowledge and Learning*, Senior Editor of *Journal of Applied Research in Higher Education*, and Deputy Editor-in-Chief of *International Journal of Learning Analytics and Artificial Intelligence for Education*.

Pedro Isaías is an associate professor at the Information Systems and Technology Management School of the University of New South Wales (UNSW—Sydney), Australia. Previously he was an associate professor at the University of Queensland, Brisbane, Australia. Before moving to Australia, he was also an associate professor at the Universidade Aberta (Portuguese Open University) in Lisbon, Portugal, where he was responsible for several courses and director of the master's degree program in Management/MBA. Pedro was director of the master's degree program in Electronic Commerce and Internet for 10 years. He is co-founder and president of IADIS—International Association for Development of the Information Society, a scientific nonprofit association. He holds a PhD in Information Management (in the specialty of information and decision systems) from the New University of Lisbon. Author of several books, book chapters, papers, and research reports, all in the information systems area, he has headed several conferences and workshops within the mentioned area. He has also been responsible for the scientific coordination of several EU-funded research projects. He is also a member of the editorial board of several journals and program committee member of several conferences and workshops. Currently he conducts research activity related to MIS in general, and more specifically Learning Technologies, Data Analytics, Business Intelligence, Digital Transformation, e-Business, and WWW-related areas.

Demetrios G. Sampson is a Professor of Digital Systems for Learning and Education at the Department of Digital Systems, University of Piraeus, Greece, where he serves as Academic Staff since 2001, and a Professor of Learning Technologies at the School of Education, Curtin University, Australia, since 2015. He is the coauthor of 340 articles in scientific books, journals, and conferences, and the editor of 12 books, 32 special issues, and 35 international conference proceedings. He has received ten times Best Paper Award in international conferences. He has been a keynote/invited speaker/lecturer in 85 international/national conferences and/or postgraduate programs around the world. He has been project director, principal investigator, and/or research consultant in 70 research and innovation projects with external funding at the range of 16 M€. He has supervised 155 honors and postgraduate students to successful completion. He has developed and delivers the first MOOC on the use of Educational Data Analytics by School Teachers (Analytics for the Classroom Teacher), offered by the edX platform which has attracted more than 12,000 participants from 160 countries around the world since October 2016. He served as Editor-in-Chief of one of the first open-access journals in educational technology, the *Educational Technology & Society Journal*, 2003–2018. He has also served or serves as member of the Steering Committee and/or Advisory and/or Editorial Board of 25 international/national journals, in various leadership roles in 75 international conferences, and at the program committee of 525 international/national conferences. He is the recipient of the *IEEE Computer Society Distinguished Service Award* (July 2012) and named a *Golden Core Member* of IEEE Computer Society in recognition of his contribution to the field of Learning Technologies (January 2013). He is also the recipient of the *Golden Nikola Tesla Chain Award* of the International Society for Engineering Pedagogy (IGIP) for "*International outstanding achievements in the field of Engineering Pedagogy*" (September 2018).

About the Authors

Kousuke Abe holds a master's in engineering. He completed his master's program at the Course of Information and Computer Sciences, Graduate School of Engineering, Kanagawa Institute of Technology in 2021. He engaged in research and development of lecture support system while attending school. He is interested in lecture support, learning analytics, and web application technology.

Catherine A. Bacos is a PhD student in the Interaction and Media Sciences program in the Teaching and Learning Department at the University of Nevada, Las Vegas. She received her MEd in Curriculum and Instruction: Educational Technology at UNLV and her BA in English and American literature from New York University. Her research interests include examining social interactions in virtual and mixed reality environments and designing learning systems for children with autism.

Josef Buchner is academic staff and PhD candidate at the Learning Lab, University of Duisburg-Essen, Faculty of Educational Sciences. His research interests include the interplay between instructional design and educational technology. Currently, he focuses on learning with Augmented and Virtual Reality technologies. ORCID: https://orcid.org/0000-0001-7637-885X.

Michael Burkhard is a PhD Student at the Institute for Educational Management and Educational Technologies at the University of St.Gallen, Switzerland. His research interests are the digital transformation in educational management as well as artificial intelligence in education.

Li Chen is a doctoral student at the Graduate School of Human-Environment Studies, Kyushu University, Japan. She received her master's degree in education from Kyushu University in 2018. She worked as a chemistry and science teacher at Shanghai Aurora Foreign Language Middle School in China from 2000 to 2015. Her research interests include learning analytics, self-regulated learning, and collaborative problem solving in STEM education.

Le Quanda L. Cole is a PhD student in the Interaction and Media Sciences program in the Teaching and Learning Department at the University of Nevada, Las Vegas. Her research focuses on a cross-disciplinary exploration of positive identity formation within academic advising. Le Quanda's research interests include transformative social emotional learning (TSEL), identity statuses, and human–computer interaction.

Elisa Corino is associate professor in Educational Linguistics at the University of Turin. Her research interests range from language acquisition to corpus linguistics and data-driven learning. She is one of the coordinators of VALICO, the online corpus of Italian as a second language, and is active in several projects on plurilingualism, CLIL, and LSP teaching with ICT.

Pia Drews is a young doctor at the University of Rostock, Department of Neurology, Germany. During her studies (10/2013–06/2020), she accompanied the Juniorstudy—first as a student tutor and in the following as a student organizer under the direction of Prof. Dr. A. Martens. This inspired the idea of analyzing the learning success of junior students in more detail as part of her doctoral thesis. She is also interested in the research area of Communication in Healthcare.

Yanyan Duan is a master's student in the School of Educational Technology, Faculty of Education at Beijing Normal University.

Joseph P. Fiorentini is a PhD student in the Interaction and Media Sciences program in the Teaching and Learning Department at the University of Nevada, Las Vegas. His research focuses on the examination of literacy skills through video game-based stealth assessments.

Cecilia Fissore graduated in Mathematics, is a PhD student in Digital Humanities at the Universities of Turin and Genoa. Her research project concerns the development of digital methodologies for language learning, characterized by the use of a Digital Learning Environment integrated with an automatic formative assessment system. She collaborates in numerous research projects in the field of Digital Education.

Francesco Floris currently works in the Department of Molecular Biotechnology and Health Sciences at the University of Turin. Francesco's main interests are the study of learning analytics and the research of data-driven solutions for improving the student experience in a Digital Learning Environment. Francesco is a member of the DELTA—Digital Education for Learning and Teaching Advances—Research Group of the University of Turin.

Olivia Foulds is an award winning Computer and Information Science PhD student at the University of Strathclyde, Scotland. With a First Class Honors in Psychology from the Open University, Olivia's research brings a multidisciplinary approach to tackle a wide range of topics that involve visual perception. In particular, she aims to increase accessibility so that individuals are better able to access information that does not cognitively overwhelm them. Beyond improving accessibility in research, Olivia also runs peer support programs and has recently been elected as the founding chair of the Equality, Diversity, and Inclusion group for Doctoral Researchers.

Xuewang Geng is a doctoral student at the Graduate School of Human-Environment Studies, Kyushu University. He received his master's degree in education from Kyushu University in 2020. He interned and worked in Qihoo 360 and Sunny Rabbit Technology Corporation and was engaged in designing the K12 online education system from 2014 to 2016. His research interests include augmented reality, mobile learning, cognitive load, and learning analytics.

Yoshiko Goda is an associate professor at the Research Center for Instructional Systems, Kumamoto University, Japan. She used to be a director of the International Board of Standards for Training, Performance, and Instruction (IBSTPI) from 2015 to 2019. She received a PhD (Science Education) from the Florida Institute of Technology in 2004 with the partial support of a Fulbright Scholarship. Her current research interests include self-regulated e-learning, online education program evaluation, technology-enhanced language learning, and innovative communities for global education.

Josef Guggemos is Assistant Professor of Educational Technology and Computational Thinking at the Institute for Educational Management and Educational Technologies at the University of St.Gallen, Switzerland. Josef's research interests are computational thinking, social robots in education, competencies of teachers and human resource developers in the context of digital transformation, information literacy, and business education science.

Said Hadjerrouit is Professor of informatics and mathematics education at the University of Agder (UiA), Kristiansand, Norway. He got a PhD degree in 1992 in the field of artificial intelligence and a master's degree (1984) in software engineering from the Technical University of Berlin, Germany. His research at the Department of Mathematical Sciences (UiA) in the period 1996–*2010* focused on informatics education, computational thinking, programming, Web-enhanced learning, and Web engineering. His current research focuses on the interactions between mathematics education, emergent digital technologies, and learning paradigms. He is also the leader of the PhD program in mathematics education.

Nils-Kristian Hansen is an Assistant professor of informatics and mathematics education at the University of Agder (UiA), Kristiansand, Norway. He got a master's degree in electronics and computer science in 1987 and a bachelor's degree in mathematics in 1996. He has spent half his professional life in developing telecommunications software, and the other half in teaching ICT and mathematic in teacher education at the Department of Mathematical Sciences (UiA). He currently also teaches a course in programming with applications in mathematics. Hansen's research interests focus on the interplay between programming, computational thinking, and problem solving in mathematics.

Oriko Harada graduated from the Graduate School of Informatics and Engineering at the University of Electro-Communications. She and her research team at the Kashihara Laboratory focuses on new opportunities for learning using information technology. Research topics at the team include knowledge building using Web resources, improving presentation skills by using robots for reflecting the movements and para-language during the presentation, and promoting ESL students to speak more with robots to decrease the uncomfortableness of conversation.

Danielle L. Head is a PhD student in the Interaction and Media Sciences program in the Teaching and Learning Department at the University of Nevada, Las Vegas. She received her MEd in Curriculum and Instruction: Educational Technology and her BA in Secondary Education (Mathematics) at UNLV. Her research interests include the design and validation of innovative psychometric assessments.

Dirk Ifenthaler is Professor and Chair of Learning, Design and Technology at the University of Mannheim, Germany, and UNESCO Deputy Chair of Data Science in Higher Education Learning and Teaching at Curtin University, Australia. Dirk's research focuses on the intersection of cognitive psychology, educational technology, data analytics, and organizational learning. He is the Editor-in-Chief of *Technology, Knowledge and Learning*, Senior Editor of *Journal of Applied Research in Higher Education*, and Deputy Editor-in-Chief of *International Journal of Learning Analytics and Artificial Intelligence for Education*.

Akihiro Kashihara received his BS and MS degrees in Computer Science and System Engineering from Tokushima University in 1987 and 1989, respectively, and his PhD in Information and Computer Science from Osaka University in 1992. He was an assistant/associate professor of the Institute of Scientific and Industrial Research in Osaka University from 1992 to 2004. He is now a professor of Graduate School of Informatics and Engineering in the University of Electro-Communications. Since 1990, he has been engaged in knowledge engineering and knowledge-based systems especially in designing intelligent and adaptive learning environments. His main research interests lie in designing of learning models, robot in learning, metacognition in hyperspace, knowledge externalization, cognitive loading, and scaffolding with cognitive tools. He has been active in both local and international academic societies, which include executive committee member of *International Journal of Artificial Intelligence in Education* (IJAIED), executive committee member of Asia-Pacific Society for Computers in Education (APSCE), and the president of Japanese Society for Information and Systems in Education (JSiSE). He has also served as Local Steering Committee Vice Chair in ICALT2007, LOC Chair in

ICCE2014, senior program committee in AIED2007–2015, etc. in several international conferences.

Michael Kerres is a full professor of educational sciences and director of the Learning Lab. He holds the chair of Educational Technology and Knowledge Management within the Faculty of Educational Sciences of the University of Duisburg-Essen. His research focuses on instructional design and how it can be applied to shape the future of education. ORCID: https://orcid.org/0000-0002-7419-3023.

S. Kathleen Krach is an Associate Professor at Florida State University in the school psychology program. Her research focuses on the development of culturally appropriate, technology-based, psychological assessments, and interventions. Dr. Krach holds her NCSP credential and is licensed as a psychologist in three states.

Martin Kursch is a member of the Department of Andragogy and Education Management at the Faculty of Education, Charles University in Prague (Czech Republic). In his academic and publishing activities, he focuses on the area of use of ICT in education and in the area of talent management. He is the author of a number of professional publications and papers in the field of education, ICT, and adult management. He has large experiences in international management and leading projects.

Jeffrey R. Laferriere is an Assistant Professor and Director of Secondary Education at the Lebanon Valley College. He specializes in the development of innovative approaches to training pre-service teachers using technology and games.

Sam A. Leif is a PhD student in the Interaction and Media Sciences program in the Teaching and Learning Department at the University of Nevada, Las Vegas. Their research focuses on the development and assessment of expertise in video game players. They also work as a research assistant in the Interaction and Media Sciences lab which specializes in human-computer interaction.

Xinyuan Li is a master's student in the School of Educational Technology, Faculty of Education at Beijing Normal University.

Tong Lu is a master's student in the School of Educational Technology, Faculty of Education at Beijing Normal University.

Marina Marchisio is full professor of Complementary Mathematics at the Department of Molecular Biotechnology and Health Sciences of the University of Turin and Delegate of the Rector for the development and promotion of the

University's Digital Education. Her research is focused on learning and teaching Mathematics and STEM disciplines with digital methodologies and innovative technologies. She coordinates the DELTA—Digital Education for Learning and Teaching Advances—Research Group. She is a member of the Problem Posing and Solving working group of the Italian Ministry of Education. She coordinates several research and didactic projects, held and organized numerous conferences, and is the author of many publications on Digital Education.

Alke Martens is Professor and Chair of Practical and Didactics of Informatics at the Institute of Computer Science at the University of Rostock, Germany. For close to 10 years she is the leader of the Juniorstudium Project, which has been founded by Prof. Tavangarian. Alke's research interests are in the intersection of AI in learning, psychology, ethics, and teacher education for computer science. She is also Chief Digitalization Officer of the University of Rostock and, among many other activities, co-founder and part of the ASLERD (Association for Smart Learning Ecosystem and Regional Development) group.

Kazunori Matsumoto is Professor at the Department of Information and Computer Sciences, Faculty of Information Technology, Kanagawa Institute of Technology, Japan. Kazunori's research focuses on data mining and artificial intelligence. He holds a PhD in engineering and is a member of IEEE, ACM, IEEJ, IPSJ, JSSST, and JSAI.

Michael P. McCreery is an Associate Professor and Director of the Interaction and Media Sciences Lab at the University of Nevada Las Vegas. His research focuses on examining how principles of human-computer interaction can be applied to the creation of game-based psychoeducational assessments and interventions.

Lucie Paulovcakova is a member of the Department of Andragogy and Education Management at the Faculty of Education, Charles University in Prague (Czech Republic). In the academic environment, she has long specialized in management in education with a focus on quality management, strategic management, marketing management, and personal development management. In these areas, she publishes in the form of scientific publications or professional contributions within scientific conferences, is a member of teams of projects with professional topics in the field of education, and performs teaching activities. She has several years of experience in the position of vice-rector of a university, where she applied a managerial approach within the management of an educational institution.

Peter Pfänder is academic staff and technology team leader at the Learning Lab, University of Duisburg-Essen, Faculty of Educational Sciences. His research focuses on the question how the role of an instructor or educator is affected or changed by educational technology.

Sergio Rabellino is a research technician at the Department of Computer Science, University of Turin. He co-operates with research groups in different topics, from Eidomatics, High Performance Computing, Artificial Intelligence to E-learning. He is a Moodle Developer and hardware/software architect of HPC computers and e-learning platforms. Technical head and architect of the Moodle-based projects Start@Unito, Orient@mente, iLearn and at the University, he counts more than 40 publications about e-learning tools and methods.

Matteo Sacchet currently works at the Department of Molecular Biotechnology and Health Sciences at the University of Turin. Matteo's main interest is teaching and learning Mathematics and, in general, STEM disciplines using digital technologies and proper related methodologies, integrated in a Digital Learning Environment. Now, Matteo is a postdoctoral researcher focusing on e-learning projects and open online courses. He is a member of the DELTA—Digital Education for Learning and Teaching Advances—Research Group of the University of Turin.

Muhittin Sahin is an academic staff at Ege University, Turkey, and postdoctoral researcher at the University of Mannheim, Germany. His research interests relate to instructional technology, learning analytics, educational data mining, multi-criteria decision making, statistic, sequential analysis, e-assessment, and intelligent learning management systems. He also collaborates in a project about developing an adaptive intelligent tutoring system.

Sabine Seufert is Director of the Institute for Educational Management and Educational Technologies at the University of St.Gallen, Switzerland. Sabine's research focuses on digital transformation in educational management and human resource development, artificial intelligence, and social robots in education.

Atsushi Shimada received his ME and DE degrees from Kyushu University in 2007. He is a professor in the Faculty of Information Science and Electrical Engineering, Kyushu University. He also worked as a JST-PRESTO researcher from 2015 to 2019. His current research interests include learning analytics, pattern recognition, media processing, and image processing. He received the MIRU Interactive Presentation Award (2011, 2017), MIRU Demonstration Award (2015), Background Models Challenge—The First Place (2012), PRMU research award (2013), SBM-RGBD Challenge—The First Place (2017), ITS Symposium Best Poster Award (2018), JST PRESTO Interest Poster Award (2019), IPSJ/IEEE-Computer Society Young Computer Researcher Award (2019), and CELDA Best Paper Award (2019).

Stefan Sonderegger is Research and Teaching Assistant at the Institute for Educational Management and Educational Technologies at the University of St.Gallen, Switzerland. His research focuses on social robots and artificial intelligence in education and learning analytics.

Zuzana Svobodova is a member of the Department of Andragogy and Education Management at the Faculty of Education, Charles University in Prague (Czech Republic). In her academic and publishing activities she focuses on the area of management of the pedagogical process and people management. She is the author of a number of professional publications and papers in the field of education management. She has many years of experience in leading positions in education and participates in domestic and foreign projects.

Tetsuo Tanaka is Professor at the Department of Information and Computer Sciences, Faculty of Information Technology, Kanagawa Institute of Technology, Japan. Tetsuo's research focuses on software engineering, web application technology, lecture support system, and learning analytics. He holds a PhD in information science; he is a fellow of IEEJ and a member of ACM, IEEE CS, IPSJ, IEICE, and JSSST.

Tobias Thelen is deputy managing director of the Center for Digital Teaching, Campus Management and Higher Education Didactics at Osnabrueck University and member of the Artificial Intelligence group at the university's Institute of Cognitive Science. He is working on the joint development of Open Source software for higher education and is interested in fostering development cooperation between higher education institutions. Tobias' research focusses on recommender systems for learning and analyzer component for adaptive learning systems.

Michaela Tureckiova is Deputy Head of the Department of Andragogy and Educational Management at the Faculty of Education at Charles University, Prague (Czech Republic). In her academic and publishing activities she focuses on the area of continuing education and professional development of employees, educational and people management, competence-based management/development and leadership. She is the author or coauthor of several scientific publications and research papers in the fields. She leads or participates on domestic and international research projects in the area of adult education and people management. She is the member of councils of experts of People Management Forum and Association of Adult Education Institutions and member of Czech Andragogy Society.

Jaroslav Veteska is Head of the Department of Andragogy and Educational Management at the Faculty of Education at Charles University, Prague (Czech Republic). In his academic and publishing activities, he focuses on the area of adult and senior education, theory of andragogy, cognitive psychology, continuing education and professional development of employees, educational management, and competence-based management. He is the Editor-in-Chief of the book edition *Czech and Slovak Andragogy*.

Anne-Cathrin Vonarx is academic staff at the Learning Lab, University of Duisburg-Essen, Faculty of Educational Sciences. Her interests focus on instructional design and educational technology in higher education and lifelong learning.

Hiroyuki Watanabe is a technical staff in the Faculty of Arts and Science at Kyushu University. He received his master's degree in Law from Seinan Gakuin University in 1998 and Graduate School of Instructional Systems from Kumamoto University in 2013. He worked as an administrative staff at Seinan Gakuin University, Japan, 1981–2019. His research interests include study skills, learning assistance, student tutors, and learning analytics.

Felix Weber is a research fellow at the Center for Digital Teaching, Campus Management and Higher Education Didactics at Osnabrueck University and a PhD student at the Institute of Cognitive Science at Osnabrueck University. He is interested in building research cooperations on topics related to digital assistants for the individualization of studies in higher education. His research focusses on the human-machine interaction with AI-powered data-driven assistants and suitable algorithms and technologies.

Masanori Yamada is an Associate Professor in the Faculty of arts and science, and the Learning Analytics Center at Kyushu University. He is engaged in learning analytics research, in particular, the relationship between learning behaviors and psychological factors such as self-regulated learning. He worked for NTT Comware Corporation where he was engaged in research on the technology related to moving images from 2000 to 2003. He received his MA and PhD in Human System Science from the Tokyo Institute of Technology in 2005 and 2008, respectively. He is the recipient of several best paper awards such as ICWL 2015 and CELDA 2016.

Guoqing Zhao is an associate professor in the School of Educational Technology, Faculty of Education at Beijing Normal University. His research interests include teaching thinking, scientific inquiry learning, and PLC-based teachers' professional development.

Part I
Learning Environments Focused on Schools

Chapter 1
Adaptive Exercises and Formative Assessment for English Remedial Action

Elisa Corino, Cecilia Fissore, and Marina Marchisio

1.1 Introduction

According to the annual report of the EF EPI (Education First—English Proficiency Index) which detects the level of knowledge of the English language, the level of fluency in English by Italians is among the lowest in Europe (EF EPI, 2019). This situation clashes with the globalized nature of the world and may preclude opportunities for future employment. School has a primary role in assuring a proper language education and the development of skills, which will allow students to successfully enter foreign universities or the labor market. Although the Italian national guidelines for secondary education state B2 as the exit level to be reached at the end of upper secondary school, according to data from the EF EPI, only 30% students have it. In addition, the percentage of students who reach B2 level falls from 40% in urban schools to 25% in provincial schools. These worrying results are confirmed by the INVALSI tests (Italian standardized tests) in the English discipline. INVALSI (https://invalsi-areaprove.cineca.it/) has adopted the levels established in the CEFR—Common European Framework of Reference for Languages (Council of Europe, 2018)—to evaluate the preparation of students in listening and reading. According to the INVALSI Report of 2019 (INVALSI, 2019), in the listening test, the percentage of grade 13 students who do not reach level B2 is

E. Corino · C. Fissore (✉)
Dipartimento di Lingue e Letterature Straniere e Culture Moderne, Università degli Studi di Torino, Turin, TO, Italia
e-mail: elisa.corino@unito.it; cecilia.fissore@unito.it

M. Marchisio
Dipartimento di Biotecnologie Molecolari e Scienze per la Salute, Università degli Studi di Torino, Turin, TO, Italia
e-mail: marina.marchisio@unito.it

D. Ifenthaler et al. (eds.), *Orchestration of Learning Environments in the Digital World*, Cognition and Exploratory Learning in the Digital Age,
https://doi.org/10.1007/978-3-030-90944-4_1

approximately 50% in the two macro-areas of Northern Italy, 64% in the Center, 79% in the South, and 84% in the South and Islands. In the reading test, the results are better, but the trend is the same as that observed for the listening test. At grade 13 students who do not reach level B2 are 35% in the two macro-areas of Northern Italy, 48% in the Center, 59% in the South, and 66% in the South and Islands. Despite its universally acknowledged importance and pervasiveness in our everyday life, it appears that English is one of the subjects where students have most difficulties and motivation. This might be partly due to the persistence of traditional teaching, mainly focused on theoretical grammar skills and structural knowledge over functional competence and communication, which is still widely practiced despite the guidelines coming from research and good teaching practices. To face the difficulties in English, teachers and institutes, but also students and parents, devote many resources in terms of time and actions taken.

At the same time, the need for students to develop digital skills, defined as the ability to know how to use technologies familiarly and critically, is emphasized in the reference frameworks of educational institutions of all levels. Another important aspect is to promote the use of technologies for educational purposes. As the research of the last two decades highlights, modern technologies and among them CALL (computer-assisted language learning) and MALL (mobile-assisted language learning) can offer valuable support for learning and teaching languages. Teachers to enrich lessons in the classroom, all the more to involve the least motivated or most struggling students (Alizadeh, 2016; Wu, 2003), can use them. The literature about the use of ICT (information and communications technology) in the language classroom is substantial, and there are lots of projects focusing on good practices to integrate the active use of technology with the last waves of theoretical research about language teaching and learning – "iTILT2 (Interactive Teaching in Languages with Technology)" (http://www.itilt2.eu), "CATAPULT (Computer Assisted Training and Platforms to Upskill LSP Teachers)" (http://catapult-project.eu), and "PETALL (Pan European Tasks for Language Learning)" (http://petallproject.eu/petall/index.php/en/) among others – but little has been done about the specific needs of remedial work.

In the light of that, the University of Turin has designed, studied, and developed innovative methodologies to help remedial work and reduce school failures (Corino et al., 2020). These methodologies are characterized by the use of a digital learning environment (DLE), a virtual space shared by students and teachers, both for face-to-face activities and for remote online activities. Our DLE, implemented on a Moodle platform integrated with an automatic assessment system (AAS) and a web conference service, is based on constructivist theory, student-centered learning, and collaborative learning (Miller-First & Ballard, 2017). The context of this research is an online recovery action for the English discipline for upper secondary school students, which involved 11 classes for a total of 265 students. The school is the Amaldi Sraffa Higher Education Institute of Orbassano (in the province of Turin), and students from grade 9 to grade 11 were involved. The recovery action, carried out thanks to the use of the DLE, should have included afternoon remedial activities in the classroom and remote online activities, both conducted by university tutors of the Master's Degree in Modern Languages and Literatures specially trained. However, due to the Covid-19 emergency and the consequent closing of the schools, the experience took place entirely online, and the tutors only met their students through the platform. The action lasted about 3 months (from March 2020 until the end of the school year in May 2020). During this period, the tutors held weekly synchronous online meetings, and they designed and created remedial activities and resources, always available to students on the platform.

Our research questions are: "What activities and resources were designed and implemented by the tutors for the recovery action, also considering the exclusively online nature of the experience?" and "Did the students improve their language skills and their interest in the discipline? Has confidence in the possibility of success increased?" To answer these questions, online courses and the activities and resources created by tutors were analyzed in detail. The methodologies used by the tutors for the implementation, with the support of the trainers, of the platform activities were also analyzed. Students' participation in all the activities and resources present in the various courses was analyzed. To study the improvement in their performance, the average evaluation in English given by the teacher before and after the recovery action was analyzed. Answers to an initial questionnaire (before the action) and a final questionnaire (at the end of the action) were analyzed to analyze students and teachers' satisfaction and their point of view on recovery.

1.2 An Interactive Digital Learning Environment for English Remedial Work

The development, the implementation, and the assessment of language learning and teaching activities have significantly increased with the use of online technologies and mobile devices (Milthorpe et al., 2018). ICT and language education have been a fundamental combination and a precious resource for foreign language teachers since the 1960s of the last century, when technologies started to be seen as a necessary support to language learning (Cuccurullo, 2012). Over the years, there have been different approaches to the use of technologies – often influenced by the theories of language acquisition – and numerous technological tools have been used in the teaching-learning processes of foreign languages. In recent years, CALL and MALL have contributed to forge paths to language learning in ways that do not replace the work of teachers in the classroom but complement it in significant manners. The importance of ICT as components of the teaching-learning environment is generally acknowledged by literature (among the manifold works on the integration of technology and the language classroom, Chapelle & Sauro, 2017; Stanley, 2013) in relation to all the benefits that this combination can have on the students' abilities, cognitive styles, learners' autonomy, and motivation. Yet, online, blended, e-learning, or technology-enhanced learning remains an area of significant challenge for humanities that traditionally prize face-to-face pedagogical methodologies such as the lecture, and small group discussion and debate (blended). A DLE, i.e., an online virtual space shared by teachers and students, can allow teachers to enhance the teaching and learning of foreign languages. Currently, various platforms are available (e.g., Padlet, Edmodo, GSuite, Moodle), some free and some not, some simpler and more basic and some more complex with plug-ins and possible integrations with other software. What we mean by DLE integrated into teaching is much more than a platform.

With the term DLE, we specifically mean a learning ecosystem in which to teach, learn, and develop skills in the classroom, online or in blended mode. In a DLE it is important both the technological component and the human component and how the activities are designed for the interactions between students, teachers, and peers. Through a DLE one can propose many different types of activities in a single shared environment; it is important in an online teaching context, but it can also integrate the teaching experience into ordinary teaching in classroom-based or in a blended mode. And what is more, the availability of a platform on mobile devices can render even greater benefits for both language teachers and learners (Herrera Mosquera, 2017). A DLE fosters the establishment and expansion of a community by offering a "real" context for interaction and supports virtual communities, with no physical walls defining the borders of the place where the community resides and works (Ligorio & Van Veen, 2006). From the didactic and pedagogical point of view, DLEs should be designed to support and foster constructivist learning, thus enabling students to actively build their knowledge with the mediation of experiences and relationships with the environment and community (von Glasersfeld, 1989), stimulating motivation and autonomous investigation, and redefining the role of teachers as facilitators rather than sources of knowledge. A central factor in the learning environment is feedback: frequent and well-structured feedback helps learners understand where they are, where they are going, and what they should do in order to reach their goal, giving information not only about how the task was performed but also about the process to be mastered and enabling self-regulation and self-monitoring of actions (Hattie & Timperley, 2007). Digital technologies can support and expand feedback practices, and, returning individualized information to students and teachers in real time, they can support and enhance learning processes. Among the requirements of an effective DLE that fulfils the constructivist and experiential ideal, there should be the possibility for students to create and publish their own works, immediately usable and shared within the community, to compare it, and to actively collaborate with the teacher or other students and the implementation of activities that promote active exploration for students and that react to the user's action giving a feedback. Activities in which digital tools are used demonstrate ways that technology can be deployed to engage students, to look at texts in fresh ways, and to make learning more interactive and student driven (Milthorpe et al., 2018).

1.3 The Online Action for English Remedial Work

The University of Turin (Unito) has designed, studied, and developed innovative methodologies for the reduction of school failures and has applied them in different contexts (Cavagnero et al., 2015; Corino et al., 2020; Fissore et al., 2019, 2020; Giraudo et al., 2014). The Unito DLE is a Moodle platform (managed by the ICT services of the Department of Computer Science of the University of Turin), integrated with the AAS Moebius Assessment for the creation of online formative tests. This AAS allows the creation of adaptive questions that provide students another chance when they give an incorrect response and that can be adapted to provide more information, let the student try a simpler version of the question, lead the student through the exercise one step at a time, and present whatever other approach the instructor feels is appropriate. This type of questions with interactive and immediate feedback is very suitable for automatic formative assessment (Barana et al., 2019; Marello et al., 2019). The platform (available at the link https://scuolapertutti.i-learn.unito.it/) is also integrated with a web conference system for online tutoring and for all contents the EasyReading font with high readability is used. Using Moodle in teaching and learning is not innovative. The novelty of our work consists in the use of a DLE integrated with an AAS, to carry out activities that involve students, tutors, and teachers (online or in the classroom when it is again possible) in a little explored and studied context such as that of recovery. The DLE allows to propose a remedial work intervention different from the traditional one, which in addition to the development of knowledge and skills related to learning English is also concentrated on enhancing interest in the matter and trust in the possibility of success. The remedial work action of English took place within the third edition of the "Scuola per Tutti" (which in English means "School for Everyone") project. The project took place in the secondary school Amaldi Sraffa of Orbassano in the province of Turin in order to recover school failures of 9th, 10th, and 11th grade students in English, Mathematics, Latin, and Physics (Barana et al., 2020). The goals of the project were to overcome learning difficulties, to reduce the phenomenon of early school leaving, to increase motivation in studying, and to facilitate the transition between lower and upper secondary school. In this research we will focus exclusively on the third edition of the remedial work action for English, as the first two editions have been used to design and test methodologies for teaching English. The remedial work action lasted about 3 months in the second part of the 2019/2020 school year. At the beginning of the project, the teachers selected the students with deficiencies in English (insufficiency between 4 and a half and 5 and a half). Throughout the period, students were followed by university tutors, specially trained on the project methodologies and on the use of the technologies involved. Initially the remedial work action included activities within the DLE of three types: synchronous face-to-face activities, synchronous online activities via web conference, and online asynchronous activities. Due to the emergency from Covid-19 and therefore the closure of the schools, the remedial work activities took place exclusively online remotely via the DLE. This led to a redesign and reformulation of the remedial work action. On the one hand, synchronous online meetings have been enhanced to give tutors and students the opportunity to get to know each other and relate as much as possible to each other. On the other hand, asynchronous platform activities have been enhanced to give students the opportunity to carry them out when they could and when they wanted (compatibly with daily online lessons). A further consequence was that of deciding to involve all the students in the class, in addition to the students selected by the teachers, to offer everyone support during the emergency. The teachers were involved throughout the course and played a key role: they indicated to the tutor the topics to be addressed with the students, reminded the students to connect to the online meetings and carry out the activities, and supported the tutor in creating the material. Conversely, the tutors constantly updated the teachers on the activities carried out with the students, on the results achieved, and on the involvement of the students. Finally, the tutors received constant support from the researchers: in implementing the project methodologies, in the use of technologies, in making the materials more interactive and training, and in relating with the students. The remedial work action was organized in this way: during the online meetings, the tutor reviewed the disciplinary topics indicated by the teacher, after which they created and made available material that students could explore and do during the week before the next meeting.

1.4 The Design of the DLE for the Recovery Action

The remedial work action involved 11 classes of 4 teachers for a total of 265 students. The classes were grouped into five groups, and five interactive online courses were set up, to which students, their teacher, and tutor always had access. Each course was managed by one of the three tutors (two tutors followed two courses). The use of the DLE makes it possible to track all student activities within the course.

The structure and format of the course were the same for each remedial course and are represented in Fig. 1.1. At the top of each course page has been included a forum for communications from the tutor (to remember online appointments or to indicate the didactic resources and activities available), a forum for student questions, and the link for the online meeting. Any time students could use the course forum to ask theoretical questions to the tutor, both on the available materials and on the use of the platform. Students could themselves answer questions in the forum in order to collaborate and help each other.

The course has a modular and visual format with a grid of icons (one for each section) with short titles. The contents of the sections are hidden and are displayed by clicking on one of the icons. Within each section, interactive materials and tests with automatic assessment created by the tutor on the topics covered during the online meetings were inserted. The number and type of materials are directly connected to the requests and needs of the students. Completion was automatically tracked for each activity and resource, in order to automatically monitor the activities carried out by students.

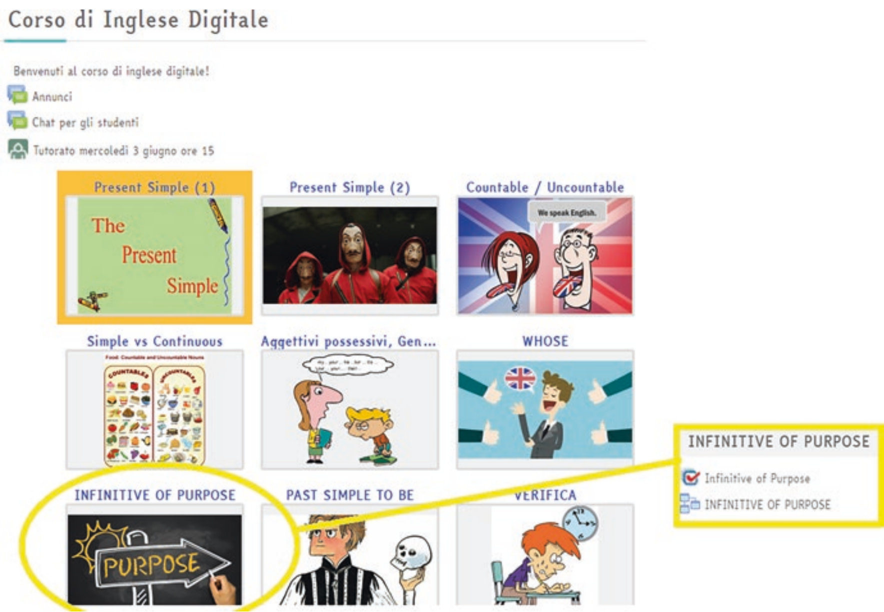


Fig. 1.1 Example of a course for recovery action

1.5 Methodologies

To analyze the activities designed and implemented by the tutors for the recovery action within the digital learning environment, the five courses on the platform were analyzed in detail. Through the creation of a configurable report, a Moodle plugin that allows you to create customized reports in SQL language has been automatically reported for each course:

- Name and id of the course
- Course teacher
- Total number of activities and resources present in the course
- Number of files (pdf, word, website, etc.) present in the course
- Number of activities present in the course for each of the following types: quiz with automatic evaluation, interactive lesson, chat, survey, glossary, workshop, questionnaire, SCORM packages, podcasts

Then the activities and resources were classified into six categories: tests with automatic assessment, interactive adaptive lessons, interactive resources, submissions, podcasts, and static resources. This classification was made both at the level of individual courses and in general at the level of action. The methodologies adopted by tutors were analyzed to make digital materials as interactive as possible, to involve students more and give them immediate and interactive feedback. Some representative examples will be shown.

To analyze the students' participation in the recovery action and in the activities and resources proposed by tutors, a second configurable report was created, to automatically analyze the number of accesses of each student to the various contents. In fact, the students could access the remedial course and the contents at any time and not only during the online tutoring with the tutor. The tutor could also implement within the course both synchronous activities, to be carried out together at the same time, and asynchronous activities that the students could carry out when they wanted and could. The report reported the following information for each student of each course:

- Student data (name, surname, email)
- Course in which he/she is enrolled
- First and last access to the course
- Number of views for files (pdf, word, website, etc.) present in the course
- Number of logs in the activities present in the course for each of the following types: quiz with automatic evaluation, interactive lesson, chat, survey, glossary, workshop, questionnaire, SCORM packages, podcast, online meeting

To analyze the students' improvement at the end of the recovery action, a t-test was carried out between the evaluation of each student before starting the recovery action and at the end of the action. The evaluations were provided by the teachers.

To understand if the students and teachers appreciated the remedial work action and if they considered it useful to recover their gaps, their answers to the final questionnaire, completed at the end of the remedial work action, were analyzed. The students' questionnaire was divided into four parts and asked them to reflect on the relationship with the English discipline (before and after the action), the remedial course and the methodologies adopted, the digital learning environment, the online mode of the action, and the technologies used. The structure of the teachers' questionnaire was the same, and in the first part, they had to evaluate the student's relationship with English from their point of view. The questionnaire was optional for the students, and, having to fill it in at the end of the school year, not everyone replied. In total 101 students replied.

1.6 Results

1.6.1 Analysis of the Activities and Resources

A total of 217 activities and resources for students on multiple topics were created in the various courses. Figure 1.2 shows a summary of the classification of the activities and resources in the various categories. At the beginning of the course, tutors shared many static resources such as theoretical pdf files, concept maps, PowerPoint presentations, images, or URLs for further information. This can be explained in several ways: at the beginning of the course, the tutors were not skilled

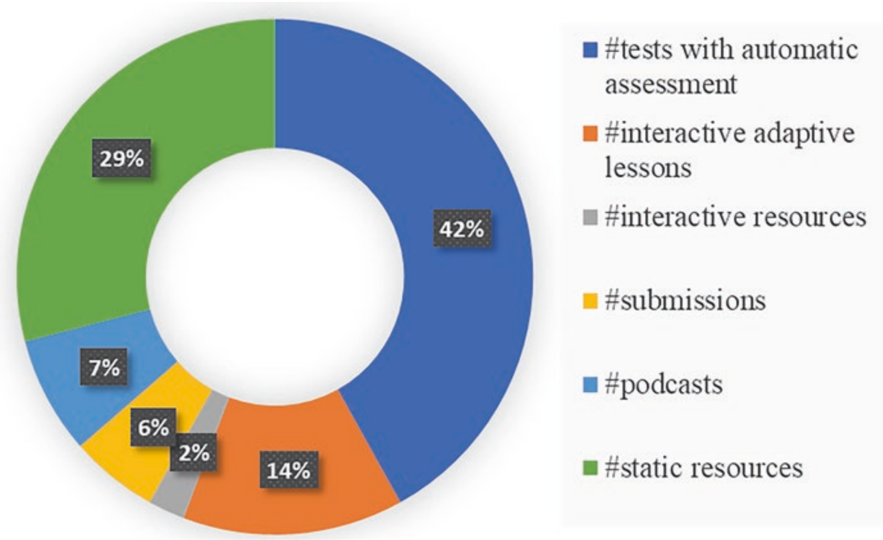


Fig. 1.2 Summary of the classification of the various activities and resources in the remedial work action

enough to exploit the potential of the DLE and chose to use tools that they were already familiar with; it is very easy to find very valid and interesting open static resources on the net; in some cases the tutors used references to textbooks that all students had. These contents can be very useful, such as videos or podcasts to practice listening or pages and conceptual maps for review; however making them available to students may not be enough. The researchers then suggested to tutors to create adaptive interactive lessons: a collection of customizable html pages. Each page can have content (video, audio, image, text, concept map, etc.) or questions. Student can navigate between the various pages and view different contents according to the choice (answer given, go to the next page, go to the previous page, select a specific page, etc.). Different comments/answers can be presented depending on the choice and students can be directed to navigation. For example, a lesson on the infinitive of purpose was proposed. In the first part of the lesson, students were asked to watch and listen to a video of a very famous English song ("The Scientist" by Coldplay) with English subtitles, in which the construct was often used. In the second part of the lesson, a summary explanation was proposed through a concept map and a narrative explanation. After this explanation, students could choose whether to watch the video again or try to take a short quiz. In the quiz different sentences were presented, and the students had to understand if the "to" in the sentence represented an infinity of purpose. For example, in the sentence "I went to the cinema to watch The Avengers!," students could choose from the answers only the first, only the second, or both. Students had personalized feedback depending on the response, for example, if they selected "only the first," they had the feedback "Pay attention! The first 'to' is a prepositions of movement"; if they selected "both," they had the feedback "Try to reason, where is the purpose in this sentence?" Students who answered all questions correctly concluded the lesson, while students who made mistakes were presented with an additional page of explanation. The lesson could also be repeated several times. This type of adaptive presentation of content and questions makes students the protagonists of their learning, forces them to think about what is required, and stimulates them to be active in building their knowledge.

Often, at the end of an interactive theory review lesson, students were offered tests with automatic assessment and immediate feedback. These tests, as shown in Fig. 1.2, represent 42% of the activities carried out. For the creation of the tests, the tutors took inspiration from exercises in the textbooks, from open material on the web, or from exercises provided by teachers. Initially tutors tended to reproduce the exercises by transcribing them with some modifications within the AAS. In the example question in Fig. 1.3, students had to think about using "some" and "any" and complete sentences correctly.

In this case, the students were assessed after having submitted the whole test, they did not get immediate feedback, and the simple "wrong answer" feedback was used, thus not providing any information about the kind of mistake and how to overcome it. It is true that any feedback would have probably been overseen by students, who hardly look at it after submitting the test and seeing the score. On top of that, they are probably no longer focused on the single item because they went through all the questions of the assignment.

Trainers guided the tutors to create adaptive questions, more effective for formative assessment. For example, the question seen above can be transformed into a question for formative assessment with immediate feedback. As in Fig. 1.4, the four cases can be proposed sequentially. After completing the first request, students can click on the "verify" button (which means "check" in Italian) and check the accuracy of the answer given, obtaining immediate feedback. In this way, if students make a mistake at the first request, they can rethink the reasoning and try to correct themselves, having experienced the previous result. After that, it is possible to insert feedback on the review of the theory directly within the question. In this way, students can actively review the rule, thinking about the answers entered and having more attempts available. By having immediate feedback, they can better acquire the general rule and put it into practice immediately. Finally, other requests are presented in the last part of the question to see if the student has internalized the rule

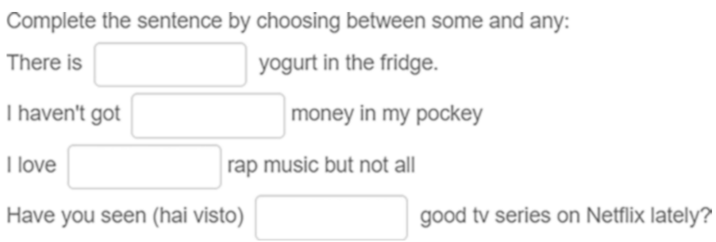


Fig. 1.3 Example of initial question created by tutors with the AAS

Fig. 1.4 Previous example transformed for a more formative assessment

correctly. The possibility of having more attempts available through a guided and interactive path is very important for the self-confidence of the students and can help to overcome the errors due to the correct insertion of the written answer. This type of activity allows students to practice independently when they prefer and to have immediate feedback on the activity carried out. Students can retry the test as many times as they wish and ask questions to the tutor or classmates in the forum if necessary. The tutor can view the students' answers and all attempts to answer, to be aware of the most frequent errors and any difficulties encountered.

During synchronous online meetings, students can compare each other on the activity carried out, and the tutor can give general feedback. In both activities – lessons and tests – adaptive teaching is proposed to activate effective formative assessment strategies, to give different feedbacks according to the level and responses of the students, and to activate self-regulated learning.

For other types of activities, such as writing a text or pronunciation exercises, tutors used the submission activity: the students submitted an online task and tutors returned feedback after a few days.

In online education it is important to give continuous feedback to students, but at the same time, it can be important to receive it from students. For example, after a phonetic activity, a tutor submitted short surveys to students in order to get feedback on the usefulness and appreciation of the activities carried out. A student wrote that she really enjoyed the activity and that the most stimulating aspect was "Knowing that there is no need to pronounce words like a native Englishman would do." This confirms that speaking in English is one of the activities in which students have the most difficulties. The latest types of activities shared by the tutors are SCORM (Shareable Content Object Reference Model) activities, a collection of specifications that enable interoperability, accessibility, and reusability of web-based learning content. All activities and resources within the DLE were accessible to students

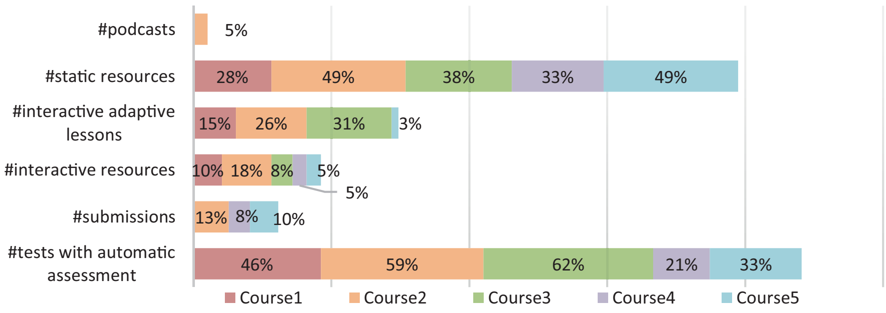


Fig. 1.5 Classification of the various activities and resources at the course level

at any time, via computer or mobile device. All students could choose when to dedicate themselves to the review activities. The activity with students gave the tutors the opportunity to receive very useful feedback and to reflect on the prepared activities.

Figure 1.5 shows the classification of activities and resources in the various categories at the course level. Course1 and course3 were managed by the same tutor, as well as course4 and course5. As can be seen from the graph, the results confirm with what emerged at a general level. The activity most used by tutors was tests with automatic assessment, followed by static resources and interactive adaptive lessons. The tutor of the course2 was the only one who used all the categories of activities and resources, in order to propose different types of activities depending on the purpose and the satisfaction of the students. Course1 and course3 are characterized by a great use of tests with automatic assessment which in both cases represent 46% and 64% of the course contents. The tutor of course4 and course5 is the one who used the least number of interactive adaptive lessons and the greatest number of static resources.

1.6.2 *Student Participation and Improvement at the End of the Action*

To analyze the participation of all students in the various courses, a configurable report has been created. The report shows for each student the number of views of activities and resources, of all the types described above. Students with zero views (17% of the sample) were excluded from this analysis. For each student, the total number of views has been calculated and the ratio between the total number of views and the total number of activities and resources present in the course, which we will call R. For example, if R is equal to 1, it means that the student has n views of n activities and resources. It is not certain that the student has viewed all the contents of the course, because he/she may have viewed the same activity n times. However, this case is unlikely, and the variable R gives a good indication of how much the student interacted with the course contents. In Table 1.1, the following have been reported for each course: the minimum value, the maximum value, the mean, the standard deviation, and the median of *R* and the division into classes of the value of *R* of all students in the course.

The minimum value of *R* represents students who have interacted very little with the course content. The maximum *R*-value is greater than 2 in almost all courses. This means that at least one student interacted for twice the course content. It is important to note that in course2 this value is higher than 4 and in course3 it is higher than 9. The average of *R* in almost all courses is close to 0.5 (so on average students interacted for half of the course contents), and in course2 it is higher than 1. As shown in Table 1.1, the value of the standard deviation indicates a lot of variability between the values of *R*. To understand how the value of *R* is distributed among the students of each course, we have decided to divide the values of *R* into five classes: 0–0.25, 0.26–0.5, 0.51–0.75, 0.76–1, and 1+. As can also be seen from the representation of the division into classes in Fig. 1.6, the number of students in the first two classes is very high (as indicated by the mean and median values). However, the graph shows that, at least for the first three courses, there is a good number of students in the last class, who have therefore participated a number of

Table 1.1 Analysis of the *R*-value of all students for each course

	R (No. of total views/No. of total activities)					Division of *R* into classes				
	Min	Max	Mean	Dev.St.	Median	0–0.25	0.26–0.5	0.51–0.75	0.76–1	1+
Course1	0.03	2.62	0.49	0.58	0.22	21	8	1	2	8
Course2	0.02	4.50	1.06	1.24	0.38	10	2	1	1	9
Course3	0.02	9.31	0.58	1.25	0.26	32	20	4	1	9
Course4	0.04	1.85	0.34	0.38	0.19	15	7	3	0	2
Course5	0.03	2.85	0.44	0.60	0.15	19	2	2	3	4

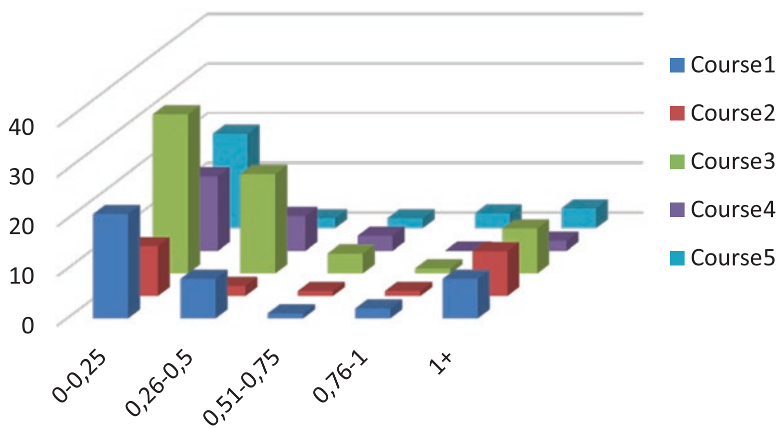


Fig. 1.6 Representation of the subdivision of *R* into classes

times higher than the course contents. This result was not found in course4 and course5. This may probably be because the tutor used content that is more static.

To study if the students improved at the end of the recovery action, a t-test was carried out for paired samples between the variables: initial evaluation (average evaluation before the action) and final evaluation (average evaluation at the end of the action). The test shows that the recovery action produced a statistically significant increase in the students' evaluations (p-value $\ll 0.05$). To investigate whether the increase in assessment is different in the various courses, we have introduced a new "increase" variable for each student, which represents the difference between final evaluation and initial evaluation. The general average increase of students who participated in the action is 0.60. This value is highest in course2 (average value of 0.78) and in course3 (average value of 0.76), where the increase is almost one grade, while the lowest value is found in course4 (average value of 0.28).

1.6.3 Students' and Teachers' Satisfaction with the Remedial Work Action

In the final questionnaire, students had to give themselves a score from "1=insufficient" to "5=excellent" in different aspects of English included in Table 1.2. The results are good because in all aspects the scores have improved (except for the participation in the class lesson, but this could be difficult to assess given the emergency). The teachers also confirm these results. They were asked to give a score to each student, with the previous scale, before and after the course. All aspects have improved. In particular, the average score increased from 2.84 to 3.35 in the knowledge of the subject, from 2.85 to 3.36 in the study at home, and from 2.88 to 3.38 in the confidence in the possibility of being able to recover. Ninety percent of teachers thought the course was useful. The most significant reasons given by the teachers regarding the success of the course were "the course proposed the discipline in a different and alternative way," "the student has improved a lot thanks to the course," "The course allowed the student to consolidate elementary contents that could no longer be reviewed in the class," and "The course changed the student's attitude to study." As regards the reasons for the failure of the course, the reasons were "the

Table 1.2 Comparison of students' answers on various aspects of English between before and after the remedial work action

	Mean before	Dev.St. before	Mean after	Dev.St. after
Interest in the subject	3.59	0.85	3.72	0.92
Knowledge of the subject	3.27	0.85	3.57	0.83
Participation in the class lesson	3.73	0.87	3.64	0.94
Study at home	3.22	0.92	3.35	0.88
Confidence in the possibility of being able to recover	3.54	1.04	3.78	0.92

student generally showed little interest in the subject, despite having acquired knowledge thanks to the course" and "In the latter part of the school year, too many commitments and too many hours spent in front of the screen prevented a profitable." Teachers noted an improvement in the attitude toward the discipline in 71% of the students. On a scale of "1 = very little" to "5 = very much," the teachers believe that the action was effective for the recovery of students in the discipline (mean 2.9 and standard deviation 0.7). They also believe that the DLE and the online tutoring allowed the students to develop new skills (mean 2.9 and standard deviation 0.8) and that the recovery project for the student was useful as a support to the normal distance learning activity (mean 3.1 and standard deviation 0.8). Students appreciated the course modalities in the following aspects (which they rated with the same scale as before): the number of meetings (3.78), the explanations of the tutor (3.65), relationship with the tutor (3.88), and online dating as a support during an emergency (3.30). The students also appreciated the fact that they have interactive materials always available on the platform with an average value of 3.57 and the interactivity of these materials with an average value of 3.41. To the question "Do you think that interactive theory and review materials can be more useful and interesting than static ones?," 81% of students answered yes. The most significant responses to the request to explain why were "interactive materials stimulate interest in the material treated," "students are more involved," "I find it useful to receive the correction immediately," "with interactive materials you apply your knowledge and understand better," "they can be more fun," and "because they stimulate you to be always concentrated and at a distance it is a necessary thing because at home we are constantly distracted." Most of the students who answered no reported that "taking lessons in class is irreplaceable" or that they prefer classical study to technology. Regarding the tests with automatic assessment, the students appreciated the following aspects indicating a score from "1= not at all" to "5=very much": perform the tests on the computer (3.53), always have access to tests with the possibility to try them several times (3.60), have immediate feedback (3.80), have multiple attempts to respond (3.41), and have questions of different types available (3.43). The students were asked "If you had difficulties in the course subject, did you manage to recover them?," with the same scale as before. The answer is very positive because the average value of the answers was 4.47 with a standard deviation of 1.47.

1.7 Conclusions

In this paper, an online action for the remedial work of English has been presented. The action was carried out online through an integrated DLE, which allows to propose student-centered learning in which students are active and responsible participants in their own learning. Students can choose between the different types of activities and resources proposed by the tutors and carry them out independently when they prefer, get feedback on the work done, and actively collaborate with the tutor or with the other students.

Our research questions concern the design and implementation of activities and resources within the DLE, their use by students, and their effectiveness in improving language skills. To answer these questions, the online courses, activities, resources, and students' participation were analyzed in detail. To study the improvement in students' performance, the average evaluation in English before and after the recovery action was analyzed. To understand if students and teachers appreciated the action and if they considered it useful, their answers to the final questionnaire, completed at the end of the remedial work action, were analyzed.

Tutors, with the support of the researchers, designed and created different types of activities and resources, especially tests with automatic assessment with immediate and interactive feedback. They also thought about how to make the didactic contents more interactive in order to involve students more in an online context. Tutors had the opportunity to have a first experience of teaching support with secondary school students, which is also useful to understand if this can represent their future work. Furthermore, during the experience, they were able to train and put into practice the innovative teaching methodologies proposed. At the end of the experience, the tutors have gained skills, and we think that this type of training could also be very useful for preservice teachers.

Students appreciated the remedial work action and the interactive character of the course. Most of the students overcame their initial difficulties with the subject. The results show that student participation in the various courses was good and that the recovery action produced a statistically significant increase in the students' evaluations. The average increase in student scores in English (before and after the recovery action) is 0.60. The opinion of the teachers was also positive. Most of the teachers felt that the course was useful and that, despite the difficulties due to the emergency period, most of the students took advantage of the experience. This remedial work action could be extended to more students and could last the whole school year in the order of giving greater and constant support. The teaching materials created by tutors could be used by the teacher for normal lessons, with some additions or modifications. In this regard, it could be interesting to enhance the training of teachers on the use of the methodologies and technologies proposed in order to use them in their daily teaching. Finally, taking inspiration from this experience, an online remedial work action for English could be designed and implemented to help students in the transition from lower to upper secondary school.

References

Alizadeh, M. (2016). The impact of motivation in English language learning. *International Journal of Research in English Education, 1*(1), 11–15.

A. Barana, F. Floris, M. Marchisio, C. Marello, M. Pulvirenti, S. Rabellino, M. Sacchet 2019). Adapting STEM automated assessment system to enhance language skills. *Proceedings of the 15th International Conference ELearning and Software for Education*. Bucharest, 403–410.

Barana, A., Pulvirenti, M., Fissore, C., & Marchisio, M. (2020). An online math path to foster the transition of students between lower and upper secondary school. In I. Carol (Ed.), *Proceedings* *of the 16th International Scientific Conference ELearning and Software for Education* (pp. 568–575). National Defence University Publishing House.

Cavagnero, S. M., Gallina, M. A., & Marchisio, M. (2015). School of tasks. Digital didactics for the recovery of scholastic failure. *Mondo Digitale, 14*, 834–843.

Chapelle, C., Sauro, S. (2017). Introduction to the handbook of technology and second language teaching and learning. In *The handbook of technology and second language teaching and learning* (pp. 1–9). Wiley-Blackwell.

Corino, E., Fissore, C., & Marchisio, M. (2020). Interactive online path for English remedial work at secondary school. In *Proceedings of the 17th International Conference on Cognition and Exploratory Learning in the Digital Age (CELDA 2020)* (pp. 181–188). IADIS Press.

Council of Europe. (2018). *Common European framework of reference for languages: Learning, teaching, assessment—Companion volume with new descriptors*. ICC.

Cuccurullo, D. (2012). La didattica delle lingue straniere con le ICT. *ANSAS, 2012*, 1–83.

EF EPI. (2019). Indice di Conoscenza dell'Inglese EF, 100 Paesi e regioni classificati in base alla padronanza dell'inglese.

Fissore, C., Floris, F., Marchisio, M., & Rabellino, S. (2019). University tutoring actions using an integrated online platform. In *Proceedings of the 6th International Conference on Cognition and Exploratory Learning in the Digital Age (CELDA 2019)* (pp. 69–76). IADIS Press.

Fissore, C., Marchisio, M., & Rabellino, S. (2020). Online tutoring through an integrated platform to support learning Mathematics at lower secondary school. In *International Conference E-Learning 2020* (pp. 71–78). IADIS Press.

Giraudo, M. T., Marchisio, M., & Pardini, C. (2014). Tutoring con le nuove tecnologie per ridurre l'insuccesso scolastico e favorire l'apprendimento della matematica nella scuola secondaria. *Mondo Digitale, 13*(51), 834–843.

Hattie, J., & Timperley, H. (2007). The Power of feedback. *Review of Educational Research, 77*(1), 81–112.

Herrera Mosquera, L. (2017). Impact of Implementing a Virtual Learning Environment (VLE) in the EFL Classroom. *Íkala Revista de Lenguaje y Cultura, 22*(3), 479–498.

INVALSI. (2019). Rapporto Nazionale Prove INVALSI 2019. https://invalsi-areaprove.cineca.it/docs/2019/Rapporto_prove_INVALSI_2019.pdf

Ligorio, M. B., & Van Veen, K. (2006). Constructing a successful cross-national virtual learning environment in primary and secondary education. *AACE Journal, 14*(2), 103–128.

Marello, C., Marchisio, M., Pulvirenti, M., & Fissore, C. (2019). Automatic assessment to enhance online dictionaries consultation skills. In *Proceedings of the 6th International Conference on Cognition and Exploratory Learning in the Digital Age (CELDA 2019)* (pp. 331–338). IADIS Press.

Miller-First, M. S., & Ballard, K. L. (2017). Constructivist teaching patterns and student interactions. *Internet Learning Journal, 6*(1), 25–32.

Milthorpe, N., Clarke, R., Fletcher, L., Moore, R., & Stark, H. (2018). Blended English: Technology-enhanced teaching and learning in English literary studies. *Arts and Humanities in Higher Education, 17*(3), 345–365.

Stanley, G. (2013). *Language learning with technology: Ideas for integrating technology in the classroom*. Cambridge University Press.

Von Glasersfeld, E. (1989). Constructivism in education. In T. Husen & T. N. Postlethwaite (Eds.), *The international encyclopedia of education* (pp. 162–163). Pergamon Press.

Wu, X. (2003). Intrinsecò motivation and young language learners: the impact of the classroom environment. *System, 31*(4), 501–517.

Chapter 2
Learning How to Use a Digital Workbench: Guided or Explorative?

Josef Buchner, Anne-Cathrin Vonarx, Peter Pfänder, and Michael Kerres

2.1 Introduction

In the digital age, more and more processes are digitized to support people in their work and professional tasks. Consequently, many new tools and applications are developed with this in mind. The challenge for developers and instructional designers is to ensure that people can use these applications and are willing to use them in the everyday work situation (e.g., Venkatesh et al., 2003). How learning to use a new digital tool or application works best is still an open question.

One tradition in educational psychology and educational research argues in favor of guided approaches that provide learners with supportive information to not overload their cognitive capacities. The effectiveness of such approaches, e.g., learning with worked-out examples or direct instruction, was demonstrated in many empirical studies (Kirschner et al., 2006; Sweller et al., 2007). However, researchers criticize that guided instruction is beneficial for structured problems and content but not for other learning objectives like affective outcomes or the invention of ideas beyond the obvious solution (Hmelo-Silver, 2004; Kapur, 2008).

Here, the other tradition comes in and argues in favor of instructional approaches with more degrees of freedom for learners. Such approaches are known as explorative learning, in which learners discover content on their own and find personalized solutions to given problems. As a consequence, motivational and affective learning objectives or learning objectives other than retention are addressed and promoted more deeply (Hmelo-Silver et al., 2007; Newman & DeCaro, 2019).

J. Buchner (✉) · A.-C. Vonarx · P. Pfänder · M. Kerres
Learning Lab, University of Duisburg-Essen, Duisburg, Germany
e-mail: josef.buchner@uni-due.de; anne-cathrin.vonarx@uni-due.de; peter.pfaender@uni-due.de; michael.kerres@uni-due.de

D. Ifenthaler et al. (eds.), *Orchestration of Learning Environments in the Digital World*, Cognition and Exploratory Learning in the Digital Age,
https://doi.org/10.1007/978-3-030-90944-4_2

In practice, most learning environments are complex and aim to facilitate different learning objectives. Hence, it is necessary to investigate which instructional approaches facilitate which learning objectives. In this study, we investigate this question and report the results of an intervention, in which participants in a workshop either learned how to use a digital workbench through a guided or an unguided instructional approach.

2.2 Theoretical Background

2.2.1 Guided Instruction

The idea of guided instruction is based on the theoretical assumptions outlined in cognitive load theory (CLT). CLT is an instructional design theory grounded on the human cognitive architecture with a limited working memory and an infinite long-term memory. Hence, the aim of instruction is to process the to be learned information in working memory and transfer it into long-term memory where it can be stored. Stored information can then be used in working memory, for example, when new tasks have to be solved (Sweller et al., 1998, 2019). The transfer of new information from working memory to long-term memory is dependent of a cognitive load imposed: high cognitive load can hinder transfer and thus learning (Sweller, 2020). As several decades in CLT research show, the learning environment, the characteristics of the learners, and the learning task influence cognitive load (Choi et al., 2014; Paas & van Merriënboer, 2020). For example, learners with little prior knowledge in a domain do not benefit from unguided instructional approaches like pure discovery or explorative learning. For them, searching for the right strategy to solve given tasks is cognitively overwhelming, time-consuming, and not better for learning performance compared to guided approaches (Kirschner et al., 2006; Sweller et al., 2007). In guided instructional approaches, learners get support from their teachers, for example, via feedback, or instructional materials like worked examples (Lazonder & Harmsen, 2016). Worked examples are elaborated samples of problems leading step by step to the solution to the problem. Novices can use them to follow an expert's approach, and as their level of expertise increases, they can detach themselves from the supportive materials (Atkinson & Renkl, 2007). Theoretically, worked examples reduce the unproductive cognitive load (extraneous cognitive load) while freeing up working memory capacities for the productive cognitive load (intrinsic cognitive load) resulting in better learning (Kalyuga & Singh, 2015).

Guiding students learning with the help of worked examples is one of the strongest CLT effects supported by plenty of empirical evidence (Sweller, 2020; Wittwer & Renkl, 2010). However, most of the CLT studies on worked examples address highly structured problems with a focus on the acquisition of domain-specific knowledge. In practice, learning environments are more complex including various learning activities with different learning objectives (Kalyuga & Singh, 2015). As a consequence, different learning objectives may need different instructional approaches. For example, exploring new solutions to a problem beyond the restrictive structure provided by worked examples might be better in learning environments with less guidance (Kapur, 2008).

2.2.2 Explorative Learning

The concept of explorative learning is a constructivist-inspired instructional approach where learners first discover content on their own followed by more teacher-led methods (Newman & DeCaro, 2019). In the literature, there are several terms describing such a procedure like productive failure or problem-solving first (Kapur, 2016; Likourezos & Kalyuga, 2017). The effectiveness of such approaches is grounded in the idea of active learning and hands-on experiences helping learners in constructing knowledge and facilitating lifelong learning skills (Hmelo-Silver, 2004). From a CLT perspective, two different assumptions with regard to explorative learning are discussed: On the one hand, explorative learning can overwhelm learners, especially beginners in a domain, because the complex tasks involved in explorative learning cognitively overload working memory capacity and thus hinder learning (Kirschner et al., 2006; Sweller et al., 2007). On the other hand, the cognitive load imposed during exploration can activate learners prior knowledge, make them aware of possible knowledge gaps, and thus boost productive cognitive processing (Kapur, 2008, 2016). For example, in Likourezos and Kalyuga (2017), learners in an unguided condition reported higher cognitive load, but no differences in learning outcomes compared to a worked examples group were found. The authors explain the result with a possible moderating effect of affect: Learners in the unguided condition accepted the more difficult task as a challenge, which in turn compensated for the higher cognitive load (Likourezos & Kalyuga, 2017, p. 214). The result is also in line with the idea presented above: different learning objectives need different instructional approaches (Koedinger et al., 2012). The promotion of interest, motivation, or attitude is a valuable aim in education, and explorative approaches of instruction might be better suited to achieve them (Newman & DeCaro, 2019). In addition, explorative learning can support students inventing new ideas and reaching learning objectives that are beyond the constructed learning scenario (Hmelo-Silver, 2004; Kapur, 2008).

2.3 The Present Study

Based on the findings above, we investigated in this study how a fully guided and an unguided explorative approach affected different learning objectives, cognitive load, and attitudes in an authentic learning environment. We situated our study within the ASPE[1] (= assessment for professional exams) project in which the team at our lab developed a so-called digital workbench. The digital workbench is an online tool based on the content management system Drupal (see Drupal.org). In the future, the workbench will support teachers and research assistants in the creation of examination tasks in vocational education. For example, a teacher can use the workbench to create text-based tasks with tables, figures, and a description of the authentic context where the tasks take place (e.g., in a transport company: cancel a purchase order because there was a problem with the delivery).

During a workshop we presented the workbench to the participants for the first time, with the possibility to test the existing functionalities and to learn how to use it. It is important to mention here that the used workbench in the workshop was a first prototype. In the meantime, the workbench has been significantly enhanced and is available in a different version.

Achieving the following learning objectives was the goal of the workshop:

- The first learning objective concerns the creation of the sub-parts for exams in vocational education. Each exam consists of several parts that, when put together, make up the final exam. Learning objective 1 was to design as many of these parts as possible. Learning objective 1 is therefore referred to as a quantitative learning objective. Since the procedure for creating the parts can be considered structured, the following first hypothesis is formulated for learning objective 1:

  The guided group will score better on learning objective 1 compared to the explorative group (H1).

- The second learning objective concerns the quality of the designed tasks. For example, every task needs a solution written in an extra text field. Again, learning objective 2 can be considered structured; thus, we test the following hypothesis 2:

  The guided group will score better on learning objective 2 compared to the explorative group (H2).

- The third learning objective concerns the functionalities of the workbench. Participants can explore a lot of functions that are beyond the traditional design of vocational exams. This task is less structured than the others; consequently the following hypothesis 3 is investigated:

  The unguided explorative group will test more functionalities of the workbench than the guided group (H3).

Furthermore, we collected data on participants' cognitive load and attitudes toward the workbench and tested the following hypotheses:

[1] ASPE is the acronym for assessment for professional exams. In this project, a digital workbench is used to support and facilitate the creation of examination tasks in vocational education and training and to set up a model of how final examinations should become more competence-based in the future.

For cognitive load, we predict in hypothesis 4 that the guided group will report lower cognitive load compared to the unguided group (H4).

For attitudes, we predict in hypothesis 5 that the guided group will report more positive and less negative attitudes toward the workbench use compared to the unguided group (H5).

2.4 Method

2.4.1 Participants and Sampling

A total of 31 participants (4 women, 22 men, 5 did not specify) with an average age of 50.6 years (SD = 8.9 years) participated in the study. The participants work as teachers and research assistants for vocational schools and a state organization for nationwide examinations in vocational education and training. The study was conducted during a workshop where the participants were able to test the workbench for the first time. With the help of colored markers on the name tag, 16 participants were randomly assigned to the fully guided worked examples group (FG) and 15 to the unguided exploratory group (UG).

2.4.2 Research Design

The independent variable in this study was the instructional approach: In the fully guided (FG) worked examples group, learners used step-by-step instructions to complete the prepared tasks. In the unguided (UG) exploratory group, participants received only a handout describing the task; no other learning materials were given. The dependent variables were the quantity (learning objective 1) as well as the quality (learning objective 2) of the completed tasks, functionalities of the workbench explored (learning objective 3), perceived cognitive load, and attitudes toward the workbench use.

2.4.3 Material

2.4.3.1 The Digital Workbench

Within the framework of the ASPE project, a digital workbench was developed to support the participants of this study in their future creation of examination tasks. In this study, the first prototype was tested for the first time. An examination in vocational education and training consists of several parts, all of which can be created in the digital workbench. For example, participants describe a so-called situational context that situates a given task; this can be a mathematical calculation, to an authentic professional context. In combination, the task and the situational context form an exam that will be presented to the students. The participants in this study had the task to create each element of an exam at least once in the same way as they would do it for a real one. The more elements created, the better for the score on learning objective 1.

In addition, each element of an exam needs certain information. For example, a solution has to be provided to every task created. If the solution is missing, the created element is of lower quality than another one. Consequently, complete elements positively influence the score on learning objective 2.

Furthermore, the workbench provides functionalities that go beyond the traditional design of vocational exams. For example, the teachers can upload tables and pictures that might help the students to understand the task. If participants used these additional functionalities, it affects the score on learning objective 3.

The prototype of the workbench allowed the research team to access all content created. The evaluation of the influence of the instructional approach on learning objectives 1–3 is based on this content.

2.4.3.2 Worked-Out Example

To guide the participants in the FG group, we designed a worked-out example with step-by-step instructions for each element of an exam. The manual consists of text-image combinations and was given to the FG group participations only in paper-based form.

2.4.4 Instruments

2.4.4.1 Learning Objectives 1–3

As mentioned above, learning objectives 1–3 were measured via the workbench. Higher values indicate higher achievement levels for all three learning objectives.

2.4.4.2 Cognitive Load

To assess cognitive load, we used the NASA Task Load Index (NASA-TLX). The NASA-TLX is a self-reporting questionnaire with six subscales (Mental Demand, Physical Demand, Temporal Demand, Performance, Effort, Frustration) answered on a scale of 0 (low) to 20 (high) (Hart, 2006; Hart & Staveland, 1988). In contrast to one-dimensional scales, the NASA-TLX is a multidimensional instrument to measure cognitive load and is therefore used frequently in cognitive load studies (de Jong, 2010; Grier, 2015). The participants answered the NASA-TLX after the intervention, *Cronbach's alpha* = .70.

2.4.4.3 Attitudes

To evaluate participants' attitudes toward the use of the workbench, we used a modified version of the personal digital assistants attitude survey (Cheng, 2017). This self-reporting survey has four subscales: Usability (four items; e.g., "The workbench is easy to use"; *Cronbach's alpha* = .90), Usefulness (four items; e.g., "The workbench supports my work"; *Cronbach's alpha* = .90), Concerns (three items; e.g., "Using the workbench was boring"; *Cronbach's alpha* = .60), and Intention to Use (three items; e.g., "I want to use the workbench in the future"; *Cronbach's alpha* = .70). The participants answered the attitude survey after the intervention on a 5-point Likert scale from 1 (do not agree at all) to 5 (totally agree).

2.4.5 Procedure

Based on the colored marker on the name tag, participants were randomly assigned to either the FG group or the UG group. For both groups separate rooms were prepared in which the participants met in small groups and worked out the tasks with the help of the workbench on notebooks. Every small group received a printed sheet with a description of the tasks to complete. Additionally, the instructional manual was given to the FG group participants. The testing of the workbench with the completion of the tasks took about 1 h. Afterward, the research instruments were answered. An overview of the research procedure is presented in Fig. 2.1. The elements created were automatically and anonymously saved by the workbench.

2.5 Results

For data analysis purpose, we first extracted all of the participants' activities from the workbench. This included data on quantity, e.g., how many of the tasks we set were completed; quality, e.g., how many of the tasks were completed in full; and exploration of functionality, e.g., which additional functions were used when creating the examination elements. All scales and subscales of the NASA-TLX and the attitude survey were aggregated using mean values. Data analysis was done in SPSS 26. The descriptive statistics for all dependent variables in this study is presented in Table 2.1. Effect sizes are interpreted according to Cohen (1992).

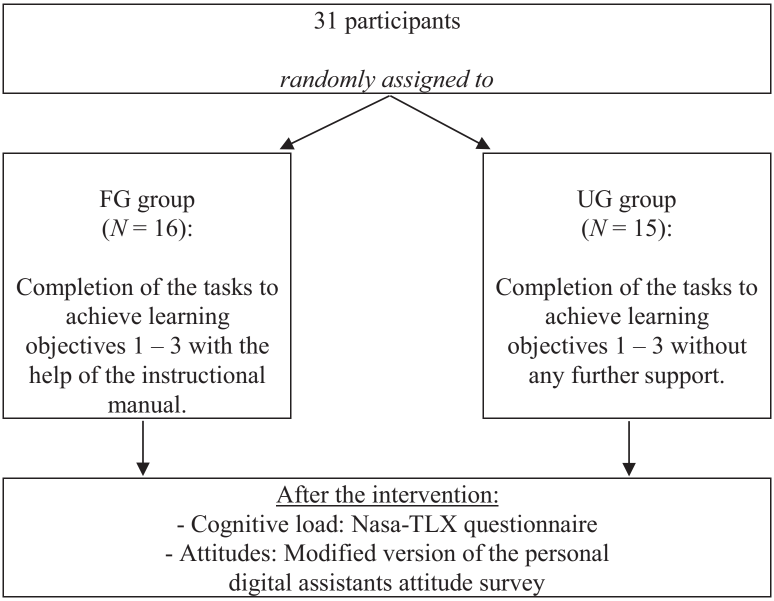


Fig. 2.1 Overview for the research procedure

2.5.1 Does the Guided Group Score Better on Learning Objective 1?

The descriptive data in Table 2.1 supports the first hypothesis as the FG group (M = 3.50, SD = 1.38) completed more tasks than the UG group (M = 1.71, SD = 0.95). To test if this difference is significant, we calculated an independent t-test. The result reveals that the FG group created significantly more elements compared to the UG group, $t(11) = -2.76$, $p < .05$, $d = 1.50$. The corresponding effect size is large.

2.5.2 Does the Guided Group Score Better on Learning Objective 2?

For the learning objective 2, the quality of the created elements in the workbench, the data in Table 2.1 supports the second hypothesis. The FG group (M = 6.50, SD = 5.68) scored higher compared to the UG group (M = 4.00, SD = 3.00). However, as the result of an independent t-test shows, the difference is not statistically significant, $t(11) = -1.02$, $p = .33$, $d = 0.37$. The effect size is small.

Table 2.1 Descriptive statistics with means and standard deviations for all variables

	FG (*n* = 16)		UG (*n* = 15)	
Variable	*M*	SD	*M*	SD
Learning Obj. 1—quantitative	3.50	1.38	1.71	0.95
Learning Obj. 2—qualitative	6.50	5.68	4.00	3.00
Learning Obj. 3—functions	8.00	5.40	10.00	6.90
Cognitive load				
Mental Demand	10.06	5.13	10.53	5.84
Physical Demand	5.12	6.12	4.40	3.23
Temporal Demand	4.00	3.76	8.92	4.73
Performance[a]	8.50	4.65	12.20	4.78
Effort	8.00	5.48	8.33	4.98
Frustration	6.25	5.62	8.80	5.67
Overall CL	6.99	3.31	8.85	3.05
Attitudes				
Usability	3.14	0.46	2.98	0.50
Usefulness	3.09	1.11	3.68	1.16
Concerns	1.46	0.47	1.58	0.56
Intention to use	3.65	0.93	3.98	0.92
Overall attitude	2.83	0.44	3.00	0.46

Note: [a]Scale performance is reverse coded. Lower values reflect a stronger belief in one's own ability to perform

2.5.3 Does the Unguided Group Score Better on Learning Objective 3?

In hypothesis 3 we predicted that the unguided group would explore more functions within the workbench. The data in Table 2.1 supports this assumption as the UG group ($M = 10.00$, SD = 6.90) scored higher than the FG group ($M = 8.00$, SD = 5.40). An independent *t*-test found that this difference is not significant, $t(11) = 0.57$, $p = .58$, $d = 0.35$. The effect size is small.

2.5.4 Does the Guided Group Report Lower Cognitive Load?

To test hypothesis 4, we calculated an independent *t*-test with the NASA-TLX subscales and the overall cognitive load as dependent variables. For the subscales Mental Demand ($p = .81$), Physical Demand ($p = .67$), Effort ($p = .86$), and Frustration ($p = .22$), no significant differences between the two groups were found. Also, the FG group and the UG group do not significantly differ in terms of the overall cognitive load ($p = .12$).

Table 2.2 Significant results of the independent *t*-test to test hypothesis 4

Variable	FG		UG				
	M	SD	*M*	SD	*t*(29)	*p*	Cohen's *d*
Temporal Demand	4.00	3.76	8.92	4.73	3.12	0.004	1.21
Performance[a]	8.50	4.65	12.20	4.78	2.18	0.037	0.78

Note: [a]Scale performance is reverse coded

Significant differences were found for the subscales Temporal Demand ($p < .01$) and Performance ($p < .05$) with large effect sizes for both (see Table 2.2). This means that the FG group perceived the completion of the task in the workbench less temporally demanding and participants were more confident in their ability to solve the tasks than the participants in the UG group.

2.5.5 Does the Guided Group Show More Positive Attitudes Toward the Workbench?

The descriptive data in Table 2.1 only partially supports hypothesis 5. For example, the FG group participants (M = 1.46, SD = 0.47) are less concerned than the UG group participants (M = 1.58, SD = 0.56). In contrast, the usefulness of the workbench is rated higher by participants of the UG group (M = 3.68, SD = 1.16) compared to the FG group (M = 3.09, SD = 1.11). In sum, the values collected are similar in both groups. An independent *t*-test confirms this impression as we found neither for one of the subscales nor for the overall attitude significant differences between the FG and the UG group.

2.6 Discussion

In this study, we compared a fully guided group with an unguided explorative group when learning how to use a new digital workbench to create vocational examination tasks. We predicted that different instructional approaches are better for different learning objectives. In hypotheses 1 and 2, we assumed benefits for the group guided through a worked-out example how to fulfil the more structured tasks given. For learning objective 1, the quantity of created tasks, the data confirms our hypothesis: the participants in the guided group significantly performed better compared to the unguided group, and the effect size is large. This result is in line with the theoretical assumptions of CLT where more guidance is beneficial for structured tasks. Furthermore, the descriptive data verifies the second hypothesis: the guided group did not only create more tasks but also with a higher quality than the control group participants. However, the difference did not reach statistical significance. A possible explanation for this is the small sample size. In a future study, a replication of the experiment is necessary with a larger sample to validate our findings. For learning objective 3, we predicted benefits for the unguided group. Here, the descriptive data confirms the assumption: the unguided group explored more functionalities in the workbench than the guided group. Again, the difference is not statistically significant, and the effect size is small. A possible explanation here might be that the participants in both groups were curious about the first use of the workbench. Hence, all participants tried to explore as much as possible in the workbench to provide feedback to the research team for the further development of the workbench. This would confirm the moderating role of affect and emotion during guided and unguided instructional approaches like mentioned in Likourezos and Kalyuga (2017).

In hypothesis 4 we predicted that the participants in the guided group report lower cognitive load than the unguided group. For the overall cognitive load and the subscales Mental Demand, Physical Demand, Effort, as well as Frustration, we found no significant differences. We found significant differences for the subscales Temporal Demand and Performance in favor of the guided group. First, this means that the participants in the guided group did not feel a strong sense of time pressure during their learning how to use the workbench. The control group without any guidance felt significantly more stressed, which can also explain the lower performance in both quantity and quality learning objectives. In addition, the guided group had more time and was therefore able to compensate the predicted lower performance with regard to the learning objective of exploring the functionalities of the workbench. More time helped the participants to create more tasks, in a high quality, and to explore as much functionalities as the control group. This is in line with CLT as guided instructional approaches are less time-consuming, better for structured learning objectives, and as good as unguided approaches for more unstructured learning objectives (Kirschner et al., 2006; Sweller et al., 2007). Second, participants in the guided group felt more self-confident to fulfill the tasks, which also resulted in a significantly better performance on learning objective 1, a better performance on learning objective 2, and almost the same good performance on learning objective 3 compared to the unguided group. This result confirms the assumption of the CLT regarding the positive effect of worked-out examples especially for beginners. As our participants saw the digital workbench in the workshop for the first time, the worked-out examples were perceived as supportive by them like shown in other CLT studies (e.g., Wittwer & Renkl, 2010).

Interestingly, we found no differences regarding the attitudes toward the use of the workbench between the two groups. Worth mentioning might be that the Usefulness and Intention to Use values are higher in the unguided group compared to the guided group. The difference is not significant, but this result needs to be further investigated by other researchers. The result is in line with the assumptions of explorative learning, whereas more freedom can address learning objectives beyond the traditional ones. Here, we found a tendency that the more affective learning objectives of Usefulness and Intention to Use the workbench in the future are slightly better addressed through exploration than guidance.

2.7 Limitations and Future Research

A limitation of this study is the small sample not allowing us to draw general conclusion about the investigated topic. The experiment also took place under real-world conditions, which is good for the ecological validity, but a further study should replicate the findings in a laboratory study. As mentioned before, future studies should also further investigate the influence of the two instructional approaches on affective learning outcomes. For example, in teacher training it would be helpful to know which instructional approach leads to more positive attitudes toward the use of a specific tool or application. Other studies should also include more modern forms of worked-out examples, e.g., video tutorials or virtual simulations of how to perform certain tasks. Here, instead of the learning how to use a digital workbench, also how to use and implement more contemporary educational technologies like augmented and virtual reality into teaching need further investigation.

2.8 Conclusion

In conclusion, the study shows that a fully guided instructional approach with worked-out examples was better able to teach participants how to use the digital workbench to solve given tasks. The guided group performed better regarding the quantity and the quality of the tasks and similar in terms of a more explorative learning objective. Furthermore, the guided instructional approach was less time-consuming, and the participants felt more self-confident when guided by the worked-out example. For the more affective learning objective of attitudes toward the use of the workbench, we were not able to find significant differences. However, we found a tendency that supports the idea of explorative learning, according to which more freedom might better address emotional and affective learning outcomes. Here, more research is needed with other forms of worked-out examples for learning how to use contemporary (educational) technologies.

Acknowledgments The authors thank all participants for taking part in the study. The ASPE project is funded by the Federal Ministry of Education and Research, funding code: 21AP007AA.

References

Atkinson, R. K., & Renkl, A. (2007). Interactive example-based learning environments: using interactive elements to encourage effective processing of worked examples. *Educational Psychology Review, 19*(3), 375–386. https://doi.org/10.1007/s10648-007-9055-2

Cheng, K.-H. (2017). Reading an augmented reality book: An exploration of learners' cognitive load, motivation, and attitudes. *Australasian Journal of Educational Technology, 33*(4), 53–69. https://doi.org/10.14742/ajet.2820

Choi, H.-H., van Merriënboer, J. J. G., & Paas, F. (2014). Effects of the physical environment on cognitive load and learning: towards a new model of cognitive load. *Educational Psychology Review, 26*(2), 225–244. https://doi.org/10.1007/s10648-014-9262-6

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*(1), 155–159. https://www.ime.usp.br/~abe/lista/pdfn45sGokvRe.pdf

Grier, R. A. (2015). How high is high? A meta-analysis of NASA-TLX global workload scores. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 59*(1), 1727–1731. https://doi.org/10.1177/1541931215591373

Hart, S. G. (2006). NASA-Task Load Index (NASA-TLX); 20 years later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 50*, 904–908.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Advances in Psychology, 52*, 139–183. https://doi.org/10.1016/S0166-4115(08)62386-9

Hmelo-Silver, C. E. (2004). Problem-based learning: What and how do students learn? *Educational Psychology Review, 16*(3), 235–266. https://doi.org/10.1023/B:EDPR.0000034022.16470.f3

Hmelo-Silver, C. E., Duncan, R. G., & Chinn, C. A. (2007). Scaffolding and achievement in problem-based and inquiry learning: A response to Kirschner, Sweller, and Clark (2006). *Educational Psychologist, 42*(2), 99–107. https://doi.org/10.1080/00461520701263368

de Jong, T. (2010). Cognitive load theory, educational research, and instructional design: Some food for thought. *Instructional Science, 38*(2), 105–134. https://doi.org/10.1007/s11251-009-9110-0

Kalyuga, S., & Singh, A.-M. (2015). Rethinking the boundaries of cognitive load theory in complex learning. *Educational Psychology Review, 28*(4), 831–852. https://doi.org/10.1007/s10648-015-9352-0

Kapur, M. (2008). Productive failure. *Cognition and Instruction, 26*(3), 307–313. https://doi.org/10.1080/07370000802212669

Kapur, M. (2016). Examining productive failure, productive success, unproductive failure, and unproductive success in learning. *Educational Psychologist, 51*(2), 289–299. https://doi.org/10.1080/00461520.2016.1155457

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist, 41*(2), 75–86. https://doi.org/10.1207/s15326985ep4102_1

Koedinger, K. R., Corbett, A. T., & Perfetti, C. (2012). The knowledge-learning-instruction framework: Bridging the science-practice chasm to enhance robust student learning. *Cognitive Science, 36*(5), 757–798. https://doi.org/10.1111/j.1551-6709.2012.01245.x

Lazonder, A. W., & Harmsen, R. (2016). Meta-analysis of inquiry-based learning: Effects of guidance. *Review of Educational Research, 86*(3), 681–718. https://doi.org/10.3102/0034654315627366

Likourezos, V., & Kalyuga, S. (2017). Instruction-first and problem-solving-first approaches: Alternative pathways to learning complex tasks. *Instructional Science, 45*(2), 195–219. https://doi.org/10.1007/s11251-016-9399-4

Newman, P. M., & DeCaro, M. S. (2019). Learning by exploring: How much guidance is optimal? *Learning and Instruction, 62*, 49–63. https://doi.org/10.1016/j.learninstruc.2019.05.005

Paas, F., & van Merriënboer, J. J. G. (2020). Cognitive-load theory: Methods to manage working memory load in the learning of complex tasks. *Current Directions in Psychological Science, 29*(4), 394–398. https://doi.org/10.1177/0963721420922183

Sweller, J. (2020). Cognitive load theory and educational technology. *Educational Technology Research and Development, 68*(1), 1–16. https://doi.org/10.1007/s11423-019-09701-3

Sweller, J., Van Merriënboer, J., & Paas, F. G. W. C. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10*(3), 251–296.

Sweller, J., Kirschner, P. A., & Clark, R. E. (2007). Why minimally guided teaching techniques do not work: A reply to commentaries. *Educational Psychologist, 42*(2), 115–121. https://doi.org/10.1080/00461520701263426

Sweller, J., van Merriënboer, J., & Paas, F. G. W. C. (2019). Cognitive architecture and instructional design: 20 years later. *Educational Psychology Review, 31*(2), 261–292. https://doi.org/10.1007/s10648-019-09465-5

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified. *MIS Quarterly, 27*(3), 425–478.

Wittwer, J., & Renkl, A. (2010). How effective are instructional explanations in example-based learning? A meta-analytic review. *Educational Psychology Review, 22*(4), 393–409. https://doi.org/10.1007/s10648-010-9136-5

Chapter 3
Longitudinal Co-teaching Projects: Scoping Review

Jaroslav Veteska, Martin Kursch, Zuzana Svobodova, Michaela Tureckiova, and Lucie Paulovcakova

3.1 Introduction

Co-teaching as an innovative method is being given increasing attention in the international environment, especially through realized experiments. Several expert studies are being conducted with research-based knowledge. Research topics are specifically focused on other forms of professional development of teachers in relation to specific educational needs, on the development of social interactions, active learning, reflective thinking, cooperation, and joint learning, with the aim of effective learning of pupils and acquisition of competences needed for active civic, professional, and personal life.

Co-teaching is defined by Bacharach et al. (2010) as joint planning and teaching of two or more teachers. The co-teaching model, where the teacher-mentor and the starting teacher perform together, uses strategies requiring shared authority, consistent involvement of both teachers, cooperation in teaching planning, and evaluation of teaching. Co-teaching, as stated in, e.g., Friend (2014, 2015) was originally based on cooperation between teachers of special and general education. Practical applications are mentioned by, e.g., Ricci and Fingon (2017). However, in the course of co-teaching, research has revealed other benefits of such guided teaching (Friend, 2015; Ricci & Fingon, 2017; Sanchez, 2019; Walsh, 2012). When applying co-teaching, it is possible, for example, to divide students into groups according to their abilities and to take advantage of the opportunity to teach students in different ways,

J. Veteska · M. Kursch · Z. Svobodova (✉) · M. Tureckiova · L. Paulovcakova
Charles University, Prague, Czech Republic
e-mail: jaroslav.veteska@pedf.cuni.cz; martin.kursch@pedf.cuni.cz; zuzana.svobodova@pedf.cuni.cz; michaela.tureckiova@pedf.cuni.cz; lucie.paulovcakova@pedf.cuni.cz; https://www.kamv.cz/; https://www.kamv.cz/; https://www.kamv.cz/; https://www.kamv.cz/; https://www.kamv.cz/

D. Ifenthaler et al. (eds.), *Orchestration of Learning Environments in the Digital World*, Cognition and Exploratory Learning in the Digital Age,
https://doi.org/10.1007/978-3-030-90944-4_3

apply different methods, and individualize teaching. The mentoring model makes it possible to improve cooperation in school environments and promote mutual learning among teachers (Baeten & Simons, 2014; Fraser & Watson, 2013; Rabin, 2020). In the framework of research into the cooperation of teachers in general education in co-teaching, better student learning outcomes, knowledge consolidation (Eckardt & Giouroukakis, 2018; Ronfeldt et al., 2015; Vescio et al., 2008). Experiments or case studies also focus on the effects of virtual co-teaching, e-learning (Chan et al., 2012; Puttonen, 2014; Takala & Wickman, 2019; Wilson & VanBerschot, 2014).

The key to co-teaching success, as Rexroat-Frazier and Chamberlin (2019) report, is the clearly defined purpose of joint teaching and the selection of partners in teaching, co-teacher education. The effectiveness of co-teaching is further influenced by teaching practices, together with the use of effective teaching strategies, effective practices, and their balance. The research carried out shows that they focus mainly on teaching practices.

Before the actual scoping review based on the analysis of the state of knowledge of the co-teaching method, we will clearly describe its individual forms and possibilities of application in teaching.

The aim of the scoping review is to analyze and summarize the most important knowledge on selected topics of co-teaching as a starting point for further research work within the project TAČR (TL03000133) entitled New Method of Education for the 21st Century: Virtual-Co-Teaching solved in the period 2020–2023. This project focuses on virtual co-teaching and its effectiveness. Therefore, also our research question asked is: "What role does virtual environment play in co-teaching?"

3.2 Co-teaching

Co-teaching is according to Beninghof (2012) defined as a "coordinated instructional practice in which two or more educators simultaneously work with a heterogeneous group of students in a general education classroom." In this definition we see three key words, which are connected to co-teaching, and those are *coordinated, educator,* and *simultaneously.*

Beninghof (2012) understands *coordination* as mutual cooperation of educators in all parts of the educational process—from planning through shared responsibility for the process to joint evaluation. The author offers a parable using synchronized swimming – sensitive, well-coordinated, and careful cooperation in a team. The term *educator* is defined very broadly by Beninghof, as all people involved in working with students in cooperative education (teachers, therapists, library staff, technology specialists, etc.). The last key word of the definition is *simultaneously*, which means that both educators are involved in the lesson.

In pedagogical practice, it is possible to identify several different models of cooperation between two educators, which define their roles and degree of involvement. For example, Murawski (2017) distinguishes between five basic types of co-teaching lessons, and Beninghof (2012) even nine. Their differences and specifics are based on three basic aspects: (1) the aspect of the roles of individual educators, (2) the aspect of the division of students, and (3) the aspect of specific educational needs. The individual types of co-teaching according to Murawski (2017) and Beninghof (2012) are presented in detail in Table 3.1.

Table 3.1 Models of co-teaching

Aspect	Type of co-teaching according to Murawski (2017)	Type of co-teaching according to Beninghof (2012)
The division of the roles of educators	1. *Team teaching* Both teachers lead the lesson; they work together	1. *Duet* Both teachers share/work together throughout the *teaching process*
		2. *Lead and support* The teaching process is comprehensively planned only by teacher A. Teacher B then participates in individual lessons
	2. *One teach—one support* The first teacher leads the pedagogical process in the lesson; the second materially supports him in individual activities	3. *Speak and add* Teacher A leads the lesson and teacher B adds to it
		4. *Learning style* Teachers plan their lessons together; *they mutually share their area of activity*. For example, teacher A is responsible for one form of teaching, and teacher B for the other
		5. *Complementary skills* Similar to "learning style." Teacher A is responsible leading and completing the lesson. Teacher B fills in based on his requirements
Division of groups of students	3. *Parallel teaching* Each teacher devotes himself to a certain part (group) of the class. A group of students can be established in different forms: a simple half, based on different criteria	6. *Skill groups* Teachers divide students into *several groups with different levels of knowledge/skills* and then treat them separately based on their level
		7. *Station* Teacher A leads the whole group/class. Teacher B deals only with a small group of students
	4. *Station teaching* Students are divided into heterogeneous groups; students change groups, and teachers deal with them individually or in groups	8. *Parallel* Teachers divide the class into two groups. *Teacher A leads one group,* and *teacher B leads the other*
Working with a specific group	5. *Alternative teaching* One teacher is dedicated to the main part of the class and the other to a smaller group that needs special care (e.g., bonus tutoring, explanation of the subject)	9. *Adapting* Teacher A leads the lesson. Teacher B *supports students who need a different approach or pace* in their lesson work

A similar approach to the division of co-teaching is described by other authors, for example, Conderman et al. (2009) and Wobak and Schnelzer (2015). Authors further supplement the individual types of co-teaching with exemplary examples of teaching activities and describe the path to quality and effective co-teaching lessons from planning and implementation to reflection.

In an *even distribution of roles*, i.e., in the "duet," both teachers share and participate in all aspects of the pedagogical process. The basic philosophy of this form of teaching is the common perception of students. Students are perceived as "our" students, not as "mine" or "yours." The benefit is mutual personal growth and direct and fast communication in solving possible obstacles. The detriment is the total amount of time devoted to teaching and related activities—educators must work very closely together and plan, create, and evaluate together.

Co-teaching in an *uneven distribution of roles* is suitable, for example, for teaching that involves a specialist together with a teacher, who normally works with the class. The teacher, who is a part of regular lessons, comprehensively prepares a given lesson, and the second teacher—specialist—works only in predetermined parts (it is then the lead and support type). The benefit of this method is the time saved for the preparation—the teacher defines the time and topics for the specialist, and there is no need to work together on the whole process. The detriment is the partial involvement of the other teacher in the training and the risk of mutual misunderstanding. Interactions with students that the practitioner is not familiar with can also be complicated.

Beninghof (2012) considers the speak and add type to be the simplest form of co-teaching. This form allows the other teacher to participate in the lesson, to be in the classroom, and to participate in the pedagogical process in the form of support in specified parts for the head teacher and the students. It is a time-saving form of preparation, the second teacher does not need to plan intensively, the lesson is in the competence of the head teacher, and the co-teaching colleague can quickly integrate. This type is also suitable for the inclusion of specialists in lessons. The use of specific approaches is made possible by the learning styles variant, i.e., again by the variant of uneven distribution of the teachers' roles.

Co-teaching colleagues work with different forms of teaching, which are adapted to the different needs of students in the classroom. For example, the first teacher is engaged in regular frontal teaching for a certain group of the students, and the second teacher works in a different way with a smaller group of students. In this form, mutual disturbance must be avoided so as not to devalue to whole lesson. A variant of this form is also the division of roles, in which one teacher focuses mainly on the explanatory part of the lesson and the other on the practical application. The complementary skills type also works on a similar basis, where the second educator (often a specialist) conducts his part of the lesson. When dividing roles, the leading role of one teacher is taken into account, and he will give the other teacher basically a "precise order," at that moment he should enter the lesson, and what and why will be a part of his section.

If we look at co-teaching in the aspect of dividing students into groups during teaching, we see the possibility of simpler internal differentiation, more effective work with a heterogeneous group, and the possibility of individualizing and developing the potential of individual students. Beninghof (2012) understands the skill groups type of co-teaching as a powerful tool that allows you to differentiate the level of the subject. Teachers divide the class into groups according to their level of knowledge and then pay them attention separately. Beninghof (2012) states that this type allows teachers to focus on developing relationships and an atmosphere in a group of students. A risk and detriment can be an unbalanced division of students or an inappropriately formed group.

The complexity of this way of working is mainly time related because dividing and planning work with separate groups requires a lot of time from teachers if they want to individualize. The form of stops allows a group of students, who have shortcomings, separate into different groups—the so-called park them at the stop and with it to focus with the help of the second teacher on problems that bother them and work with them. In this style of work, Beninghof (2012) draws attention to the significant risk posed by the stigmatization of the group as a weaker part of the class.

Students can also be divided into two smaller groups each with its own teacher. In this case, it is a so-called parallel form of co-teaching. This form is basically very simple and not very difficult for cooperation, as each teacher teaches his group, using his style, and decides his lesson. However, it is not about cooperation in the true sense of the word, but rather about dividing students into different groups for the possibility of more intensive work. When dividing students into groups, a type of co-teaching called adaptation or alternative teaching is also used, in which a special pedagogue joins the lesson and deals with students with special educational needs. The common teacher teaches normally, and the co-teaching colleague is then in charge of effective inclusive strategies and the involvement of students, who need a specific approach. This type of co-teaching is often used in inclusive schools in Germany or Austria, where it began to spread significantly after 2009, when special schools began to be abolished and students were integrated into mainstream schools. Governments then decided on the so-called double cast of lessons, which allowed schools to apply this style of co-teaching and use its opportunities for effective inclusion and work with heterogeneous groups (Wobak & Schnelzer, 2015).

The individual described types of co-teaching are usually not applied separately, but instead intertwine causing new combinations to occur. If educators decide in terms of dividing students on the skill groups co-teaching type, then it is necessary to decide in what form in terms of the roles of educators these skill groups will be led. In this case, a duet can take place, but it is also possible to divide the roles unevenly and use skill groups only in the teaching part and work using the lead and support or the speak and add type. Similarly, it is possible to consider the aspect of educational needs. Thus, the interconnectedness and continuity of co-teaching types follow from above. It is therefore up to the educators to set the goal of their teaching and choose the most appropriate type. Beninghof (2012) adds that it is not necessary to adhere exactly to the defined form, but it can be suitably adapted to the needs of students and teachers and thus get the most out of it.

3.3 Scoping Review

3.3.1 Methods

3.3.1.1 Scoping Review Methodology

We chose scoping review as the research method. This method is suitable for the systematic search, selection, analysis, and subsequent synthesis of the obtained information so that we can answer research questions responsibly. The methodological basis for our study was the work of Arksey and O'Malley (2005), which we followed in all steps of the research. Arksey and O'Malley (2005) describe the scoping review methodology as a five-step process involving (1) identifying the research question, (2) identifying relevant studies, (3) study selection, (4) charting the data, and (5) collecting, summarizing, and reporting the results. Arksey and O'Malley (2005) also define the basic purposes of scoping review, which are (1) to examine the extent, range, and nature of research activity, (2) to determine the value for undertaking a full systematic review, (3) to summarize and disseminate research findings, and (4) to identify research gaps in the existing literature.

3.3.1.2 Identifying Relevant Studies

When searching for studies, we first used the keyword "co-teaching." The result was more than 2000 studies. That's why we've added an advanced search that's consistent with our goal. Another string was the word "long-term," the results suddenly from 12 studies. We replaced the word long-term with "longitudinal," and the result was another 17 studies, different from the previous combination. Another combination was the search for "co-teaching" and virtual at the same time. The result was another ten studies. We then used the same criteria in Google Scholar. Five papers were found using the keyword "virtual co-teaching" was found. Next, we searched in the title of the article virtual and co-teaching and found five articles.

Sixteen databases (ERIC; Complementary Index; Academic Search Ultimate; Scopus®; Supplemental Index; APA PsycInfo; Social Sciences Citation Index; Directory of Open Access Journals; Gale eBooks; Springer Nature Journals; ScienceDirect; MEDLINE; JSTOR Journals; Library, Information Science & Technology Abstracts; Business Source Ultimate; Humanities Source Ultimate) and Google Scholar were searching to find as many relevant studies as possible. We limited your search to 1990–2020. The search tool was the system of Charles University UKAZ and separately the search engine Google Scholar. In Table 3.2, all the results are presented.[1]

[1] Reference to the UKAZ system, which has access to all of the above databases: http://eds.b.ebscohost.com/eds/search/advanced?vid=7&sid=b259143b-4e01-42db-8c7c-22f94f9d499f%40pdc-v-sessmgr04

Table 3.2 Exact search string results in indexed databases

Databases	Search string	Records identified
ERIC; Complementary Index; Academic Search Ultimate; Scopus®; Supplemental Index; APA PsycInfo; Social Sciences Citation Index; Directory of Open Access Journals; Gale eBooks; Springer Nature Journals; ScienceDirect; MEDLINE; JSTOR Journals; Library, Information Science & Technology Abstracts; Business Source Ultimate; Humanities Source Ultimate	co-teaching	2162
	co-teaching AND virtual	10
	co-teaching AND long-term	12
	co-teaching AND longitudinal	17
Google Scholar	"virtual co-teaching"	5
Google Scholar	allintitle:: virtual co-teaching	5

3.3.1.3 Inclusion and Exclusion Criteria

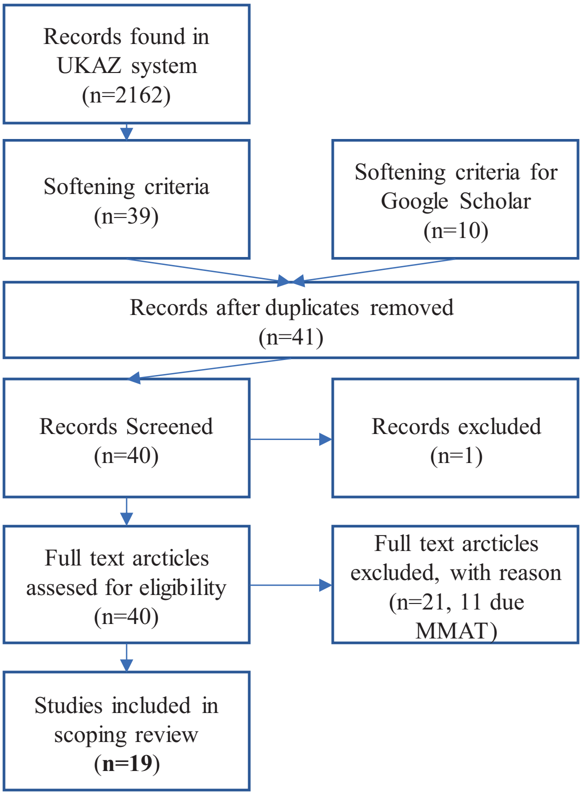
We identified a total of 49 studies that we had to go through and manually sort. The inclusion and exclusion criteria were discussed by all authors together. We focused mainly on the period 1990–2020, as previous studies are not relevant to the virtual dimension. In the end, we selected 19 studies in our sample that meet the main criterion of long-term research and deal with the effects of co-teaching or its effectiveness. We have also included all studies that study the virtual possibility of co-teaching. PRISMA systematic evaluation (Moher et al., 2009) was used, and our algorithm and results are described on Fig. 3.1.

3.3.1.4 Critical Appraisal

When evaluating the quality of studies, we used the Mixed Method Appraisal Tool (MMAT), version 2018. According to this methodology, we were able to disable studies based on proven quality criteria for mixed research and study methodology. The MMAT follow-up was used independently by two authors from our group to promote independence and eliminate subjective concepts. Based on MMAT, 11 studies were excluded.[2]

[2] See http://mixedmethodsappraisaltoolpublic.pbworks.com/w/file/fetch/127916259/MMAT_2018_criteria-manual_2018-08-01_ENG.pdf.

Fig. 3.1 Search flowchart



3.3.1.5 Data Analyses

The method of inductive thematic analysis was used for data analysis (Corbin & Strauss, 2008). The articles were sorted by content into individual categories, the authors identified key topics in them, and then the articles were regrouped according to selected key topics. In the final phase, the sub-themes were grouped according to their context by mutual agreement between the authors.

3.3.2 Results

A total of 19 articles were analyzed. The most frequently declared aim in these articles was "to analyze the experience with co-teaching" (n = 5), "describe the impact and benefits of co-teaching for students" (n = 4), "identify specifics and patterns of cooperation of teachers in the co-teaching process" (n = 4), "describe the advantages and disadvantages of co-teaching" (n = 2), "describe specific possibilities of use" (n = 2), "identify the strengths and weaknesses of the method and modify it" (n = 1), and "describe the attitudes of teachers" (n = 1). Ten of these 19 studies used a qualitative design, 3 used a quantitative design, and 6 studies used mixed methods. Most studies relied on multiple methods of data collection ($n = 12$). The most common method of collecting qualitative research data was interviews ($n = 7$). This method was supplemented by written reflection of respondents ($n = 6$), observation ($n = 4$), and open question questionnaire ($n = 2$). Statistical methods were mainly based on questionnaires and scaling ($n = 10$). An overview of the methods used can be found in Table 3.3.

Table 3.3 Summary of individual studies

Author/Year	Aim	Research type	Study design and methods	Study partic[1]ipants	Themes	Virtual
Chan et al. (2012)	Based on a three-month measurement, describe the positive impact of co-teaching on students, teachers and school principals	Mixed methods	Structured questionnaire with scale, personal interview after the lesson	3 full classes (A with leading teacher, B with supportive tacher and C observing)	Problems and obstacles, teachers' cooperation, co-teaching effectiveness, teachers' roles and relationships	Yes
Eckardt and Giouroukakis (2018)	Describe the impact of co-teaching in combination with situational learning on the teaching students	Qualitative	Analysis of anonymous reflection of students taught by co-teaching, observation during lessons and teacher's notes	21 teachers taking a co-teaching-led "teen literacy" course	Teachers' cooperation, effectiveness of co-teaching, combination of methods	No
Rabin (2020)	Describe the process of implementing co-teaching in the co-work of a mentor and a novice teacher	Qualitative	Interviews, observations, qualitative analysis of co-teaching implemenatation	241 participants, 70 mentors and 171 novice teachers	Teachers' cooperation, ethical dimension, co-teaching effectiveness, teachers' roles	No
Jurkowski and Müller (2018)	Describe the development of the relationship of a co-teaching couple composed of a special educator and an ordinary teacher	Quantitative	Questionnaire, 3 measurements/ year	36 ordinary teachers, 19 special educators, 184 students	Teacher cooperation, teacher relationships, problems and obstacles, co-teaching effectiveness, special educational needs	No
The Park (2014)	Identify specific patterns of cooperation of teachers in class	Qualitative	Analysis of 5 videos from English lessons in one class	2 teachers (native and non-native speaker)	Teachers' cooperation, teachers' roles and relationship	No

(continued)

Table 3.3 (continued)

Author/Year	Aim	Research type	Study design and methods	Study partic[1]ipants	Themes	Virtual
Neifeald and Nissim (2019)	Identify the benefits of the "Academia Class" programme and its impact on co-teaching practices	Quantitative	Questionnaire, statistical evaluation	Participants in the Academia Class project, 125 (51 teachers, 26 teachers students, 18 primary school teachers, 20 early childhood teachers, 10 trainers)	Teachers' cooperation, combination of methods, teachers relationship, co-teaching effectiveness	Yes
Ricci and Fingon (2017)	Based on the analysis of experience with co-teaching to create the basis for effective application of co-teaching at primary and secondary schools in teaching a class with pupils with SEN	Mixed methods	Evaluation and reflection of co-teaching training plans, analysis of experience; mathematical and statistical methods of data evaluation	34 postgraduate students, 14 special pedagogy, 20 literature	Special educational needs, teachers' cooperation	No
Campbell et al. (2018)	Describe the pros and cons of co-teaching by looking at university students	Mixed methods	Structured questionnaire (online, Likert scale), qualitative analysis of open responses	Not stated	Teachers' cooperation, problems and obstacles	No
Puttonen (2014)	Analyse new experience of students with online teaching led by two experts	Quantitative	Structured questionnaire (online)	32 graduates of virtual co-teaching course	Co-teaching effectiveness, teachers' cooperation	Yes
Takala and Wickman (2019)	Describe the experience of rural teachers with co-teaching and identify problems in its application	Mixed methods	Questionnaire, content analysis of open answers	40 teachers from one municipality (29 ordinary, 11 special educators)	Cooperation of teachers, problems and obstacles to co-teaching	No

(continued)

Table 3.3 (continued)

Author/Year	Aim	Research type	Study design and methods	Study participants	Themes	Virtual
Sanchez (2019)	Analyse the pedagogical experience with a postgraduate course led by the co-teaching method in order to identify the strengths and weaknesses of this method	Qualitative	Case study, written reflection of participants and its content analysis	3 cooperating experts (university teacher), 2 directors	Teachers' cooperation, problems and obstacles	No
Strogilos and King-Sears (2019)	Describe the experience and progress of students with and without SEN in co-teaching	Qualitative	Semi-structured interviews	3 students with SEN, 7 students without SEN	Co-teaching efficiency, special educational needs	No
Wilson and VanBerschot (2014)	Analyse and subsequently modify the e-learning course for master's students	Qualitative	Case study, design thinking	2 teachers and course participants	Teachers' cooperation, co-teaching effectiveness	Yes
Thompson and Dow (2017)	Describe the effectiveness of co-teaching of two experts from different disciplines compared to teaching led by one expert	Mixed methods	Action research, observation and measurement using a scale	24 science students	Co-teaching effectiveness, combination of methods, teachers' cooperation	No
Kim et al. (2007)	Describe the possibilities of using the CACSR computer program in co-teaching lessons to promote reading among students with SEN	Qualitative	Evaluation and reflection of the method	2 teams of teachers (ordinary teacher and special educator)	Special educational needs, teachers' cooperation, teachers effectiveness, teachers' roles	Yes

(continued)

Table 3.3 (continued)

Author/Year	Aim	Research type	Study design and methods	Study partic[1]ipants	Themes	Virtual
Duran et al. (2020)	Analysis of attitudes of teachers towards co-teaching in the context of conceptual training	Mixed methods	Questionnaire, analysis of written reports	82 wells as co-teachers in two groups (with and without training)	Teachers' cooperation, co-teaching effectiveness, problems and obstacles	No
Lohmus et al (2019)	Describe the benefits of co-teaching and identify its potential risks	Qualitative	Structured interviews	2 co-teachers. 4 students reflecting co-teaching at higher education	Teachers' roles, co-teaching effectiveness, teachers' cooperation	No
Montgomery and Akerson (2019)	Analyze co-teachers' collaboration	Qualitative	Reflection of experience	2 co-teachers reflecting on their practical teaching experience	Teachers' cooperation, teachers' roles, co-teaching effectiveness	No
Bilican et al. (2020)	Analyse the use of co-teaching as a method of further teacher education	Qualitative	VNOS-C questionnaire, analysis of recordings of educational activities, reflective interviews, participating observations	2 co-teachers (science teacher and university teacher)	Teachers' cooperation, teachers' roles, co-teaching effectiveness	No

3.3.3 Themes

Based on thematic analysis, six themes mentioned by the respondents in the articles reviewed were identified in the studies: (1) *co-teaching effectiveness*, (2) *problems and obstacles*, (3) *methods of and with co-teaching*, (4) *teachers' role and relationships*, (5) *teachers' cooperation*, and (6) *special educational needs*. Individual topics cannot be completely separated from each other, because, for example, the topic of co-teaching effectiveness is mingled with virtually all the texts analyzed and is always placed in another context related to another identified topic. Otherwise, effectiveness will be seen in combination with special educational needs and otherwise in the context of teacher cooperation, their roles, and relationships. It is therefore clear that in the co-teaching process, individual topics are logically intertwined and always accentuate a certain aspect related to the specific context in which this method is implemented. The following section describes the identified topics in more detail, including the relationship or context with other co-teaching topics.

3.3.3.1 Co-teaching Effectiveness

The effectiveness of co-teaching is a central identified topic, which in a way contains all the texts. It includes, on the one hand, a view of effectiveness by the teacher (e.g., Wilson & VanBerschot, 2014; Jurkowski & Müller, 2018; Neifeald & Nissim, 2019; Rabin, 2020) as well as a student view (e.g., Puttonen, 2014; Strogilos & King-Sears, 2019). The view of effectiveness by teachers is usually the basis for revision of the methods used or modification of a new co-teaching course. Based on a pilot study, the teachers reflect the entire teaching process, finding the positives and negatives on which it is based when adjusting the course or educational activity to its final form (Kim et al., 2007; Wilson & VanBerschot, 2014; Neifeald & Nissim, 2019).

In view of the research question asked, "What role does virtual environment play in co-teaching?," the technical readiness is essential in terms of efficiency (Chan et al., 2012; Wilson & VanBerschot, 2014). This factor lies (Chan et al., 2012) on the border between efficiency and problems, because with technical complications it is impossible to teach effectively with two people, the benefit of the method decreases, and complications and negative impact on results dominate. This aspect is also essential from the point of view of students who perceive the readiness of a co-teaching pair (whether technical or human).

3.3.3.2 Problems and Obstacles

Problems and obstacles of co-teaching can be divided according to three identified potentially problematic factors—the factor of technical problems, the human factor, and the lack of time—based on the texts analyzed. Technical problems are complicated mainly by co-teaching using a virtual environment. Chan et al. (2012) stated that the well-managed technical aspect of the whole process is essential for virtual co-teaching and therefore, after a pilot analysis of virtual learning, they made the necessary technical adjustments. For example, the problem of face imaging and face reading had to be technically solved by installing additional audio/video technology and cameras on the move (specifically, mobile cameras on teachers' heads). Thanks to this effect, students in the classroom were able to synchronize and at the same time include a peer learning method, which was not possible without additional technical installation. Subsequent results have shown that if virtual co-teaching is technically handled well, it can lead to positive impacts on students (Chan et al., 2012).

The human factor is the most fundamental. If cooperation does not work, teachers do not plan and evaluate joint teaching, there are significant shortcomings, and the method does not have the necessary benefits (Jurkowski & Müller, 2018). A specific area of the human factor is the approach of the school management to co-teaching. Takala and Wickman (2019) have identified the role of the headmaster as a significant obstacle in the implementation of co-teaching, since its implementation is directly dependent on the decision of the headmaster and not on the teachers, which logically leads to the fact that the headmaster who is not inclined to it will not support and implement it at the school.

It is clear that teaching one group of students with two people is challenging and necessarily requires time to prepare lessons, create a meaningful plan, split roles, set up cooperation, and then reflect on teaching and translate it into further planning and other lessons (Ricci & Fingon, 2017). However, teachers do not have enough time available (Park, 2014; Jurkowski & Müller, 2018) and therefore cannot devote themselves to preparation and reflection to the extent necessary. The lack of time factor was reported by most studies as the most important and most affecting overall effectiveness of co-teaching (Takala & Wickman, 2019; Jurkowski & Müller, 2018; Ricci & Fingon, 2017; Sanchez, 2019).

3.3.3.3 Methods of and with Co-teaching

Specific chosen methods, procedures, and forms are different in different implementations. In the analyzed texts, we encountered the following forms of co-teaching methods:

- One teach—one support, also supporting model (Park, 2014; Chan et al., 2012; Eckardt & Giouroukakis, 2018; Takala & Wickman, 2019; Lõhmus et al., 2019; Montgomery & Akerson, 2019; Duran et al., 2020; Rabin, 2020)
- Alternative teaching, also complementary or complementary model (Ricci & Fingon, 2017; Strogilos & King-Sears, 2019; Jurkowski & Müller, 2018; Kim et al., 2007)
- Team teaching (Wilson & VanBerschot, 2014; Puttonen, 2014; Thompson & Dow, 2017; Campbell et al., 2018; Neifeald & Nissim, 2019; Sanchez, 2019; Bilican et al., 2020)

In the "one teach—one support" model, in addition to standard assistance to the head teacher, it is also a method of education for future or novice teachers. In fact, the head teacher becomes a mentor at the same time, listening and learning from each other based on a partnership, which is not entirely possible with classical mentoring (Montgomery & Akerson, 2019; Rabin, 2020). Bilican et al. (2020) show co-teaching as a possible way of further education. A team consisting of a university teacher and an ordinary teacher confirms that the professional development of teachers can be effectively based on a co-teaching strategy.

Furthermore, co-teaching was supplemented by other methods, which are used as standard even in classical teaching by one teacher. Eckardt and Giouroukakis (2018) combine co-teaching with situational learning, resulting in more effective knowledge consolidation than teaching with one teacher. Chan et al. (2012) combine co-teaching in a virtual environment with peer learning by students, which increases the effectiveness of the positive impacts of co-teaching on students. The virtual part is involved in co-teaching in different ranges. It allows you to teach using e-learning courses (Wilson & VanBerschot, 2014), teach together and be different (Puttonen, 2014; Chan et al., 2012), or take advantage of two teachers and digital technologies and special educational programs (Kim et al., 2007; Neifeald & Nissim, 2019).

3.3.3.4 Teachers' Roles and Relationships

The division of roles is closely related to the chosen co-teaching method. As far as the "one teach—one support" model is concerned, it makes sense that the teaching will always have its head teacher (e.g., Lõhmus et al., 2019). Roles between the different co-teaching pairs can also result naturally from the education and approbation of individual teachers. It is logical that otherwise there will be divided roles in the teaching of pupils with special educational needs, where special educator and ordinary teacher teaches (Strogilos & King-Sears, 2019; Jurkowski & Müller, 2018; Kim et al., 2007), and otherwise in foreign language teaching, where a native speaker and teacher speaking the student's mother tongue is involved (Park, 2014). When using the virtual component of co-teaching, they are also the basis for the division of roles of digital competence of individual teachers (Kim et al., 2007). With regard to the various roles, it is also necessary to mention the ethical dimension of cooperation, where the balance of power, collegiality, and the pursuit of the same goal at the same level of mutual consent should be maintained in relations between co-teachers and feedback on teaching should be shared (Rabin, 2020).

3.3.3.5 Teachers' Cooperation

The cooperation of teachers is also based on individual methods of co-teaching and their role in teaching. However, regarding the effectiveness of co-teaching, its development is important. Teacher cooperation is in three stages—planning, implementation, and reflection. In this area, however, it also encounters the problem of the time factor, which is the most frequently mentioned obstacle to the quality implementation of co-teaching (Takala & Wickman, 2019; Jurkowski & Müller, 2018; Ricci & Fingon, 2017; Sanchez, 2019). In preparation there is room for more creative methods of teaching, and in implementation a faster and better response is possible, as well as space for improvisation (Eckardt & Giouroukakis, 2018). Based on joint reflection, it is possible to change and modify the work with students to make it more effective and impact on students greater (Wilson & VanBerschot, 2014; Thompson & Dow, 2017). In the preparatory phase of co-teaching, it is advisable for future co-teachers to use suitable workshops to help them set up mutual cooperation and lead to greater efficiency of the method (Rabin, 2020).

Cooperation of co-teachers can also be based on mutual replenishment, where the pair consists of experts with different specializations. The results demonstrate the effectiveness of co-teaching led by two subjects, leading to a deeper understanding of the subject (Thompson & Dow, 2017).

Quality cooperation of teachers also brings reliability to students. If there are two teachers, students have answers from two different sources, which they consider more reliable. Furthermore, this association opens a new perspective for them at work, because the differences in teacher experience make students see the problem from multiple perspectives and there is constructive discussion (Eckardt & Giouroukakis, 2018). Co-teaching enables the application of a wide range of teaching practices, including partnerships between teachers and teachers, shared planning, and evaluation of teaching. Without co-teaching, these procedures would be very difficult to implement, even completely unfeasible (Neifeald & Nissim, 2019).

3.3.3.6 Special Educational Needs (SEN)

A specific thematic area consists of special educational needs, in which co-teaching is used as a method that helps to lead teaching in the class of intact and SEN pupils. Studies show a positive experience (Strogilos & King-Sears, 2019), and pupils report that it was an effective "extra help" that allows them to feel valid members of the class (Strogilos & King-Sears, 2019). More flexibility of teachers and the ability to respond to individual needs have also been confirmed by these students (Strogilos & King-Sears, 2019).

Jurkowski and Müller (2018) offer another look at the use of co-teaching in teaching students with SEN, which focuses mainly on the cooperation of an ordinary teacher and a special teacher, which is not always easy. Cooperating teachers encounter a lack of planning time, which in turn also affects the possibility of a positive impact on students and the effectiveness of teaching. This aspect is also mingled with other mentioned topics and is the overall dominant mentioned problem in the implementation of co-teaching (Chan et al., 2012; Takala & Wickman, 2019; Sanchez, 2019).

3.4 Conclusion

The overview study focuses on co-teaching as an innovative method, which is also given increasing attention in terms of professional development of teachers and especially with regard to social interactions, active and joint learning, and, above all, the development of the effectiveness of pupils' learning and the acquisition of their necessary competences for a successful and active civic, professional, and personal life.

As already mentioned in the previous sections of this overview study, co-teaching is a topic whose importance is increasing over time, while virtual co-teaching is also part of the expert discussion. Co-teaching can be analyzed from different points of view. When working in the term of co-teaching, we follow the aspect of the division of teacher roles, which can be completely even, up to a clear leading role of one educator, who uses the other in precisely defined parts and topics of the lesson. We also focus on the aspect of dividing students, where co-teaching enables the implementation of different approaches and different ways of internal differentiation, which also allow effective work with students with special educational needs. However, the different areas of co-teaching and the prospects for its use are intertwined and linked to specific contexts (e.g., virtual environments), with the issue of co-teaching effectiveness linking them. Given the research question asked ("What role does virtual environment play in co-teaching?"), this co-teaching context is examined within the individual identified topics as and if it appears in the source documents.

Co-teaching efficiency, identified as a central and cross-cutting theme, includes both a view of effectiveness by the teacher (Wilson & VanBerschot, 2014; Jurkowski & Müller, 2018; Neifeald & Nissim, 2019; Rabin, 2020) and students (e.g., Puttonen, 2014; Strogilos & King-Sears, 2019). Technical readiness, standing on the border between the topic of efficiency and problems and obstacles, is crucial for the effectiveness of the virtual co-teaching process from the point of view of teachers. In addition to the mentioned technical readiness, the human readiness (cooperation) of the co-teaching two is also important for students in the virtual form of co-teaching, allowing students to orientate well in their teaching.

3.4.1 Application

The findings presented in this study and the lessons learned from previous research in selected six thematic areas can be used as a basic framework for the preparation, implementation, and reflection of co-teaching as an innovative method using different methodological procedures and forms of teaching, in which two teachers are always involved in different roles and evolving cooperation. For the effectiveness of co-teaching (including virtual co-teaching), it is important to deal with its entire process (planning, implementation, reflection, including the choice of co-teaching type and its specific variant of virtual co-teaching). Other contextual conditions that affect the effectiveness of (virtual) co-teaching and are elaborated in sub-topics in this study include technical readiness, a time factor for considering the specific educational needs of pupils.

3.4.2 Limitations

English articles were included in the overview study. This can affect the overall validity of the results found and their portability into local education systems, which are shaped in accordance with each country's education policy strategy (subsidiarity principle). Gray literature was not included in the review.

The presented study provides an overview of the types of co-teaching (including virtual) and the basic topics to be addressed in order to make this innovative teaching method effective. The results of the study will also be used in the next phases of the TAČR project[3].

References

Arksey, H., & O'Malley, L. (2005). Scoping studies: Towards a methodological framework. *International Journal of Social Research Methodology, 8*(1), 19–32.

Bacharach, N., Heck, T. W., & Dahlberg, K. (2010). Changing the face of student teaching through co-teaching. *Action in Teacher Education, 32*(1), 3–14.

Baeten, M., & Simons, M. (2014). Student teachers' team teaching: Models, effects, and conditions for implementation. *Teaching and Teacher Education, 41*, 92–110.

Beninghof, A. M. (2012). *Co-teaching that works: Structures and strategies for maximizing student learning*. Jossey Bass.

Bilican, K., Akerson, V., & Nargund, V. (2020). Learning by teaching: A case study of co-teaching to enhance nature of science educators, successes, and challenges. *International Journal of Science and Mathematics Education, 18*(1), 1–20.

Campbell, E. L., Reedy, A. R., Baird, M. J., & Baird, D. M. (2018). Better together: Co-teaching in undergraduate applied psychology courses. *Psychology Teaching Review, 24*(2), 3–11.

Boris Chan, Alan Lam, Louis Li, 2012. Case study: Affordances of using multi-HD technology for distanced co-teaching in Hong Kong secondary schools. In *2012 World Congress on Information and Communication Technologies, Information and Communication Technologies*, Trivadrum, 2012, pp. 16–166.

Conderman, G., Bresnahan, V., & Pedersen, T. (2009). *Purposeful co-teaching: Real cases and effective strategies*. SAGE.

Corbin, J. M., & Strauss, A. L. (2008). *Basics of qualitative research: Techniques and procedures for developing grounded theory*. SAGE.

Duran, D., Flores, M., Ribas, T., & Ribosa, J. (2020). Student teachers' perceptions and evidence of peer learning through co-teaching: Improving attitudes and willingness towards co-teaching. *European Journal of Psychology of Education, 35*, 1–16.

Eckardt, P., & Giouroukakis, V. (2018). The impact of co-teaching on pedagogical approaches and student conceptual understanding in a graduate, adolescent literacy course. *Journal for Leadership and Instruction, 17*(2), 40–43.

Fraser, J., & Watson, A. (2013). Bring student teaching into the 21st century. *Phi Delta Kappan, 94*(7), 25.

Friend, M. (2014). *Co-teach! Building and sustaining effective classroom partnerships in inclusive schools* (2nd ed.). Author.

Friend, M. (2015). Co-teaching versus apprentice teaching: An analysis of similarities and differences. *Teacher Education and Special Education, 38*(2), 79–87.

Jurkowski, S., & Müller, B. (2018). Co-teaching in inclusive classes: The development of multi-professional cooperation in teaching dyads. *Teaching and Teacher Education, 75*, 224–231.

Kim, A.-H., Woodruff, A. L., Klein, C., & Vaughn, S. (2007). Facilitating co-teaching for literacy in general education classrooms through technology: Focus on students with learning disabilities. *Reading, 22*(3), 269–291.

[3] Funding: This study was supported by Technology Agency of the Czech Republic (TAČR), Project No. TL03000133.

Lõhmusa, E., Naritsa, H., & Ugur, K. (2019). Lecturer's and assistant's views on collaborative teaching at the university. *Estonian Journal of Education, 7*(2), 231–233.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine, 6*(7), e1000097. https://doi.org/10.1371/journal.pmed.1000097

Montgomery, M., & Akerson, A. (2019). Facilitating collaboration through a co-teaching field experience. *Networks: An Online Journal for Teacher Research, 21*(1), 1–21.

Murawski, W. W. (2017). *Beyond co-teaching basics: A data-driven, no-fail model for continuous improvement*. ASCD.

Neifeald, E., & Nissim, Y. (2019). Co-teaching in the "Academia Class": Evaluation of advantages and frequency of practices. *International Education Studies, 12*(5), 86–98.

Park, J. (2014). English co-teaching and teacher collaboration: A micro-interactional perspective. *System, 44*(1), 34–44.

Puttonen, K. (2014). Information specialist and ICT lecturer co-teach an online course: A new way and what students think about it. *Nordic Journal of Information Literacy in Higher Education, 6*(1), 52–64.

Rabin, C. (2020). Co-teaching: Collaborative and caring teacher preparation. *Journal of Teacher Education, 71*(1), 135–147.

Rexroat-Frazier, N., & Chamberlin, S. (2019). Best practices in co-teaching mathematics with special needs students. *Journal of Research in Special Educational Needs, 19*(3), 173–183.

Ricci, L. A., & Fingon, J. C. (2017). Faculty modeling co-teaching and collaboration practices in general education and special education courses in teacher preparation programmes. *Athens Journal of Education, 4*(4), 351–362.

Ronfeldt, M., Farmer, S. O., McQueen, K., & Grissom, J. A. (2015). Teacher collaboration in instructional teams and student achievement. *American Educational Research Journal, 52*(3), 475–514.

Sanchez Jafeth, E., Kelly, H., & Carroll, K. (2019). Exploring the co-teaching experience in a graduate-level, principal preparation course. *InSight: A Journal of Scholarly Teaching, 14*, 99–112.

Strogilos, V., & King-Sears, M. (2019). Co-teaching is extra help and fun: Perspectives on co-teaching from middle school students and co-teachers. *Journal of Research in Special Educational Needs, 19*(2), 92–102.

Takala, M., & Wickman, K. (2019). Collaborative case-based virtual learning in higher education: Consultation case in special education. *Journal of Digital Learning in Teacher Education, 35*(4), 236–248.

Thompson, K. W., & Dow, M. J. (2017). Co-teaching to improve control variable experiment instruction in physical sciences education. *Electronic Journal of Science Education, 21*(5), 36–52.

Vescio, V., Ross, D., & Adams, A. (2008). A review of research on the impact of professional learning communities on teaching practice and student learning. *Teaching and Teacher Education, 24*(1), 80–91.

Walsh, J. (2012). Co-teaching as a school system strategy for continuous improvement. *Preventing School Failure, 56*, 29–36.

Wilson, B. G., & VanBerschot, J. L. (2014). Co-teaching an online action research class/ Co-enseignement et classe de recherche-action en ligne. *Canadian Journal of Learning and Teaching, 40*(2), 1–18.

Wobak, M., & Schnelzer, W. (2015). *Teamteaching: Kollegiale Kooperation für gelingendes Lehren und Lernen*. School of Education, Universität Innsbruck.

Part II
Learning Environments Focussed on Higher Education and Further Education

Chapter 4
Characterizing Personal Educational Goals: Inter-rater Agreement on a Tagset Reveals Domain-Specific Limitations of the External Perspective

Felix Weber and Tobias Thelen

4.1 Introduction

In psychology goals have been defined as "internal representations of desired states" (Vancouver & Austin, 1996). This sharp definition entails a set of properties of goals in general: At first, goals are subjective because of their internal nature. A goal always exists in the mind of a person and not externally. Even in a group of people expressing the same goal in common terms, the individual mental representations will most likely differ significantly. At second, a goal is directed toward a state, most likely in the future. This aspect holds strong implications to behavior because of the attractiveness of goal states. The third important aspect is desirability, which is subjective by definition. A person having a goal assumes, for some reason, that acting toward a future, in which the goal state becomes reality, leads to positive consequences, such as well-being. Interestingly this is not necessarily the case. Research has shown that the characteristics of goals that we chose have a much stronger impact on subjective well-being, that goal achievement itself (Ehrlich, 2012). The aspect of desirability raises a philosophical or even metaphysical question; humans have asked themselves since ancient times: What is desirable (Dallmayr, 2008)?

Goal-setting theory (Latham & Locke, 1990, 2007; Locke & Latham, 2019) postulates that by explicitly formulating personal goals, individuals are more likely to attain them. Goal-setting individuals are hypothesized to be more capable of directing their effort and attention toward goal-relevant tasks and of ignoring distractions. Indeed, the setting of goals can, in itself, bolster individuals' self-regulation capacity. Furthermore, goal setting boosts persistence, thus reducing the impact of negative influences such as anxiety, disappointment, or frustration. In addition,

F. Weber (✉) · T. Thelen
Osnabrück University, Osnabrück, Germany
e-mail: fweber@uos.de; tobias.thelen@uni-osnabrueck.de

D. Ifenthaler et al. (eds.), *Orchestration of Learning Environments in the Digital World*, Cognition and Exploratory Learning in the Digital Age,
https://doi.org/10.1007/978-3-030-90944-4_4

well-defined goals are thought to encourage individuals to develop more efficient strategies to meet their aims. Over the past 4 decades, more than 400 experimental and correlational studies have provided evidence for the validity of goal-setting theory (Locke & Latham, 2002; Latham & Locke, 2007).

Research has shown that the kind of goals humans set themselves changes over their life span. While infants approach goal states with high immediate rewards, choosing goals based on a balanced time perspective is considered to be essential for a good life in positive psychology (Boniwell & Zimbardo, 2012). At least two goal characteristics have been shown to change with increasing age: While younger people tend to be more outcome-focused, the elder tend to be more process-focused (Freund et al., 2012), and while the younger tend to have achievement goals, directed toward maximizing gains, older people tend to have maintenance goals, directed to the avoidance of losses (Freund et al., 2012). These findings make sense when young age is understood as an indicator for a life phase of growth and high age as an indicator for a life phase of decline. Thus, it can be assumed that typical university students are in a life phase of growth, with a tendency toward outcome-focused goals. Furthermore, choosing and pursuing goals based on a long-term time perspective may be challenging, especially to freshmen, because of limited experience with freedom of choice in preceding phases of education.

A fascinating emerging field in goal-setting research is the study of goal systems, which broadens the researchers' perspective from isolated goals toward a plethora of goals with various kinds of interrelations, for instance, synergistic or conflicting. There is a growing body of literature about goals and their relations (Kung & Scholer, 2018; Ballard et al., 2016; Heckhausen et al., 2010; Kruglanski et al., 2002; Weber, 2019).

Self-determination theory (SDT) by Ryan and Deci (Deci & Ryan, 1985; Ryan et al., 2017) is an organismic theory of human behavior and personality development that outlines a taxonomy of motivation and self-regulation styles. SDT postulates a continuum ranging from *amotivation*, *external motivation, introjected extrinsic motivation*, *identified extrinsic motivation*, and *integrated extrinsic motivation* to *intrinsic motivation*, where the former are controlled externally, while the latter are based on inner sources of motivation and more joyful. The process of internalization of goals in this model depends on the personal importance, self-endorsement, self-congruence, and personal interest in goals. A higher education system that aims to produce independent proactive individuals should aim to support students in the development of intrinsic motivation.

Self-regulated learning is a model of learning in which active learners play an essential role by regulating their behavior in iterative loops of *planning*, *executing*, and *reflecting learning processes* (Zimmerman, 1990). In cyclic models of self-regulated learning, individual learning goals are an essential building block because they form the foundation for planning and serve as benchmarks for evaluation (Zimmerman, 1989). Intrinsic motivation has been shown to predict the use of strategies for self-regulation and learning strategies (Virtanen et al., 2013), which underlines the importance of intrinsically motivated educational goals for personal development based on self-regulation.

There is a long tradition of applying goals to the academic arena (Morisano, 2013). Simply elaborating on personal goals and the ideal future can significantly increase academic performance (Schippers et al., 2020, Alessandri et al., 2020), especially for the extreme group of struggling students (Morisano, 2008; Morisano et al., 2010). These findings imply that students can benefit from interventions that let them think about and clarify and construct personal educational goals. This finding has significant implications on goal-setting research in higher education, as studies can be designed as goal-setting interventions with personal relevance for participants, which can potentially increase intrinsic motivation and facilitate the success of recruitment procedures.

In the German context, the EU-wide Bologna reforms have introduced new challenges, particularly for bachelor students, for whom the transition to the world of work requires an increasingly high degree of professional goal orientation and self-control (Olos et al., 2014). However, a considerable proportion of students are known to be rather poor at formulating intrinsic study goals. Studies suggest that they can greatly benefit from support to both formulate and maintain commitment to self-set goals (Schippers et al., 2020).

Theoretically the process of human goal pursuit could be modeled as a navigation through a state space toward a goal state or a set of goal states. In the material reality state, transitions are fluent and the state space is continuous and infinite. In the domain of digital higher education, where students strive toward their educational goals within a learning management system, the state space is finite, states are clearly separable, and transitions are discrete. Consequently, for goal-setting researchers, interested in digital assistant technologies, the domain higher education learning management systems is a great environment for research because it is simpler and more tractable than the real world.

The specific context in which the outlined research was done is the SIDDATA project (Thelen et al., 2018) which is funded by the German Ministry of Education and Research and aims at the development and investigation of a digital data-driven study assistant. The study assistant is intended to support students in the development and pursuit of personal educational goals. The support happens based on data from various sources and machine learning algorithms. The study assistant is supposed to support the monitoring of goal progress, which is also known to improve effective self-regulation and increase the likelihood of successful goal outcomes (Harkin et al., 2016).

Our primary goal in the research outlined here is to (a) collect a set of realistic educational goals in natural language, (b) develop a tagset which is suitable to describe the characteristics of students' goals, and finally (c) use the tagged goals as training data for machine learning models, which will then be able to automatically label unseen goals of students using the study assistant.

4.1.1 Existing Study Goal Classifications

The idea to describe the characteristics of educational goals has a long history, but a comprehensive system unifying the plethora of previous attempts is a blank spot in goal-setting research. In this section we introduce some important classification systems without claiming completeness.

Bloom's taxonomy of educational goals (Bloom et al., 1956) is an early approach to structure educational objectives. The proposed taxonomy entails the six graded levels of cognitive learning goals: knowledge, comprehension, application, analysis, synthesis, and evaluation. Since its development it has been broadly applied in higher education and is a topic of ongoing debate (Amer, 2006; Wellington, 2020).

In developmental psychology and educational psychology, there is a line of research about achievement goals (see Murayama et al. 2012) which led to a 2 × 2 model (Elliot & McGregor, 2001) and a 2 × 3 model (Elliot et al., 2011). The foundation for the 2 × 2 model is a distinction between two types of educational goals, termed *learning goals* and *performance goals* by Dweck (1986), *task-involvement goals* or *ego-involvement goals* by Ames (e.g., Ames & Archer, 1987), or *task-involvement goals* and *ego-involvement goals* (Nicholls, 1984, 2017). With *performance* (or *ego-involvement*) *goals*, students focus on their abilities and sense of self-worth, whereby ability is achieved by surpassing normative-based standards and/or the performance of others. By contrast, *learning*, *mastery*, or *task-involvement goals* reflect the belief that effort and outcome co-vary. They thus highlight intrinsic motivational patterns. In contrast to *performance goals*, *mastery goals* are evaluated on the basis of self-referenced standards (Ames, 1992).

There are some issues with theoretical clarity in motivational constructs (Pintrich, 2000a). What these conceptualizations have in common is that one conceptual extreme is mastering a task to learn or simply solve the task for its own sake, while the other conceptual extreme is solving a task to demonstrate personal competence or to perform well. These early conceptualizations converged in the *mastery goal-performance goal* dichotomy, which forms the first dimension of the 2 × 2 model. The second dimension of valence distinguishes between positive *approach goals* and negative *avoidance goals*, which was developed by Elliot and colleagues (Elliot & Harackiewicz, 1996; Elliot & Church, 1997; Elliot, 1999).

The resulting 2 × 2 matrix contains *mastery-approach goals*, *mastery-avoidance goals*, *performance-approach goals*, and *performance-avoidance goals*. Both types of *mastery goals* could be shown to correlate with intrinsic motivation, *performance-avoidance goals* correlate with dysfunctional learning strategies and low grades, and *performance-approach goals* predicted high grades (Elliot & Church, 1997).

In the 2 × 3 model, the mastery-performance dichotomy in the definition dimension is replaced by a trichotomy related to the reference frame of goal definition: A goal can be defined relative to one's own past achievements (*self*), by the concrete task (*task*), and relative to other persons (*others*). This results in six types of goals: *self-approach*, *self-avoidance*, *task-approach*, *task-avoidance*, *other-approach*, and *other-avoidance goals*.

There is empirical evidence for interaction effects these discrete variables have on outcome variables, such as intrinsic motivation, grades, and quality of learning strategies. Furthermore, there are conclusive theoretical explanations for these effects. A weak point of both of these models is simplification of discrete variables, which assumes that there are distinct sets of goals. Instead, we think it is more realistic to model the dimensions of the 2 × 2 model as continuous variables, where a goal can have components of *mastery* ("I want to understand the concept of factor analysis.") and *performance* ("I want to pass the statistics test with a good grade.") in the same goal. Furthermore, this goal characteristic may even change over time, perhaps the aspect of *mastery* is stronger when I start learning for the exam, and the aspect of *performance* reaches a peak shortly before the exam.

The 2 × 3 model is even more apart from reality, because at first, its reality is not discrete in the reference frame dimension, either, and at second, the trichotomy cannot be modeled as one continuous dimension. A prospective approach could be to develop a 2 × 2 × 2 model, perhaps with the dichotomies approach-avoidance, mastery-performance, and internal-external reference frame.

Goal-setting theory assumes that goals are cognitive representations of what individuals are trying to accomplish and their purposes or reasons for attempting a task. As such, they are inherently cognitive and assumed to be accessible by the individual. This, however, is not necessarily a given, and, in real-world contexts, students spontaneously formulate much more varied goals. Indeed, such self-set goals have been reported to "take on a much more personalized, idiographic flavor" (Elliot & Thrash, 2001). Furthermore, each achievement goal category potentially encompasses very many different sub-levels of goals. By way of illustration, Elliot and Thrash (2001) differentiate between high-level striving, e.g., "learn as much as I possibly can at school this year," and lower-level striving, e.g., "get at least 45 out of 50 problems correct on my math exam." Hence the kind of goals that students freely formulate may be simple task-based target goals, overarching goal orientations or goal complexes. In addition, goals are cognitive representations and, as such, are expected to be adapted on the basis of contextual sensibility (Pintrich, 2000b).

Although these goal categories have traditionally been theorized as dichotomous and in opposition to one another, empirical correlational studies based on survey data have reported conflicted results with some positive, negative, and nonsignificant correlations between the supposedly opposing types of goals (Pintrich, 2000a). A further issue with such goal classification schemes concerns the consciousness or cognitive accessibility of motivational constructs and thus whether students can accurately report on their own motivation (Murphy & Alexander, 2000).

For instance, Elliot and Church (Elliot & Church, 1997) devised a questionnaire to assess college students' adoption of mastery, performance-approach, and performance-avoidance achievement goals. Participants responded to six items, such as "It is important to me to do better than the other students" and "I want to learn as much as possible from this class," on a 7-point Likert scale ranging from "not at all true of me" to "very true of me." Important to note is that this questionnaire asks about the individual and not the goal, which reveals the influence of motivation research tradition.

Also in the German higher education context, Ahn et al. (2012) applied both previously established goal categories and added new categories on the basis of interviews and questionnaires with students and lecturers. These categories reflect goals related to the choice of a particular course of study. The authors classified them into six supra-categories. On the basis of the data collected, Ahn et al. (2012) proposed a four-level hierarchical model of study goals. The highest level represents the overarching aim of *living a good life*. The second level makes a temporal differentiation between goals concerning students' university studies and those concerning life after graduation. The model's third level is about the goal's contextual environment: at university vs. outside of university and private vs. professional lives. The fourth, most fine-grained level, distinguishes between self-focus vs. other-focus, high vs. low involvement in the learning process, professional vs. personal considerations, and fulfilling personal wishes vs. those of others.

As part of a personal growth goal-setting program, Travers and colleagues (Travers et al., 2015) explored the types of academic performance-related growth goals that students choose to set themselves. The authors subsumed these goals into three broad categories: (1) personal organization and time management, (2) emotional and psychological control, and (3) interpersonal-skills development, stressing that these need not be discrete categories. Similarly, as part of a large quasi-experimental goal-setting intervention program, Schippers and colleagues (2020) categorized students' self-set goals with a set of seven categories based on life domains: academic, career, social relationships, material, physical health, mental well-being, and miscellaneous. Two independent raters classified the goals according to these seven categories, and Schippers et al. (2020) report high inter-rater agreement scores of $k = 0.85–0.87$.

4.1.2 Our Approach

Previous studies on students' personal educational goals have tended to focus on one or two study disciplines (most frequently economics, management, and psychology), and attempts to classify freely formulated study goals have usually been restricted to a handful of very broad goal-type categories, with only few studies providing evidence for validity beyond the conceptual level, or reliability, such as inter-rater agreement scores.

By contrast, the present study seeks to create and validate a tagset that covers a broad range of characteristics of individual educational goals of university students. The tagset is inspired by and evaluated on the basis of a data set of goals from three German universities from a broad range of disciplines. Ultimately, it is hoped that such a tagset, together with a large reliably manually tagged training data set, may later be used by a digital study assistant to automatically analyze students' self-set goals in order to design personalized goal-directed recommendations.

4.2 Methods

In the data processing subsection, we outline the data processing starting from the recruitment of participants and ending with the final data analysis. In the following subsection about the technical setup, the software architecture, which allowed a smooth integration into the local learning management system Stud.IP[1] (Stockmann & Berg, 2005) and was mostly "hand-coded," is described. In the third subsection, the tagset is introduced and enriched with literature for further reading.

4.2.1 Data Processing

The procedure is summarized in Fig. 4.1. First, students from the universities of Bremen, Hannover, and Osnabrück across all study programs were invited to participate in the study using existing university- and faculty-specific email mailing lists, as well as advertising on the universities' local learning management system (LMS). The sampling technique used was a combination of self-selection and convenience sampling (Oates, 2006).

The link provided in the emails and the LMS adds redirected students to a web-based interface embedded in the universities' local Stud.IP learning management system. The web page detailed the SIDDATA project's wider aims of creating a digital study assistant and encouraged students to participate by citing research that has shown that the formulation of personal goals can, in itself, positively contribute to attaining them (Locke & Latham, 2002).

Further motivation was provided by a gamification element: Students were informed that, on submission of their goals, they would be able to see the n most frequently submitted goals of all students, after the data collection ended, where n is the number of goals they submitted themselves. So if a student submitted four goals, after the data acquisition ended, the student would get to know the four most frequent goals. Due to the collected goals, from which the most were unique, the most frequent goal tags were returned to students instead of the concrete goals.

In the first text box of the input interface, participants were informed of the pseudo-anonymization procedure of their personal data and could opt in to have personal information concerning their course and their current semester of study saved alongside their goals. The exact information saved was displayed next to the check box. The data collection procedure was checked by the data protection officials of the universities involved and approved to be in line with GDPR regulations.

[1] Stud.IP is an open-source campus management system (CMS) and learning management system (LMS) for universities, schools, companies, organizations, and government agencies. It provides interfaces that allow the integration of external systems and applications. More information can be found at the project homepage at https://www.studip.de.

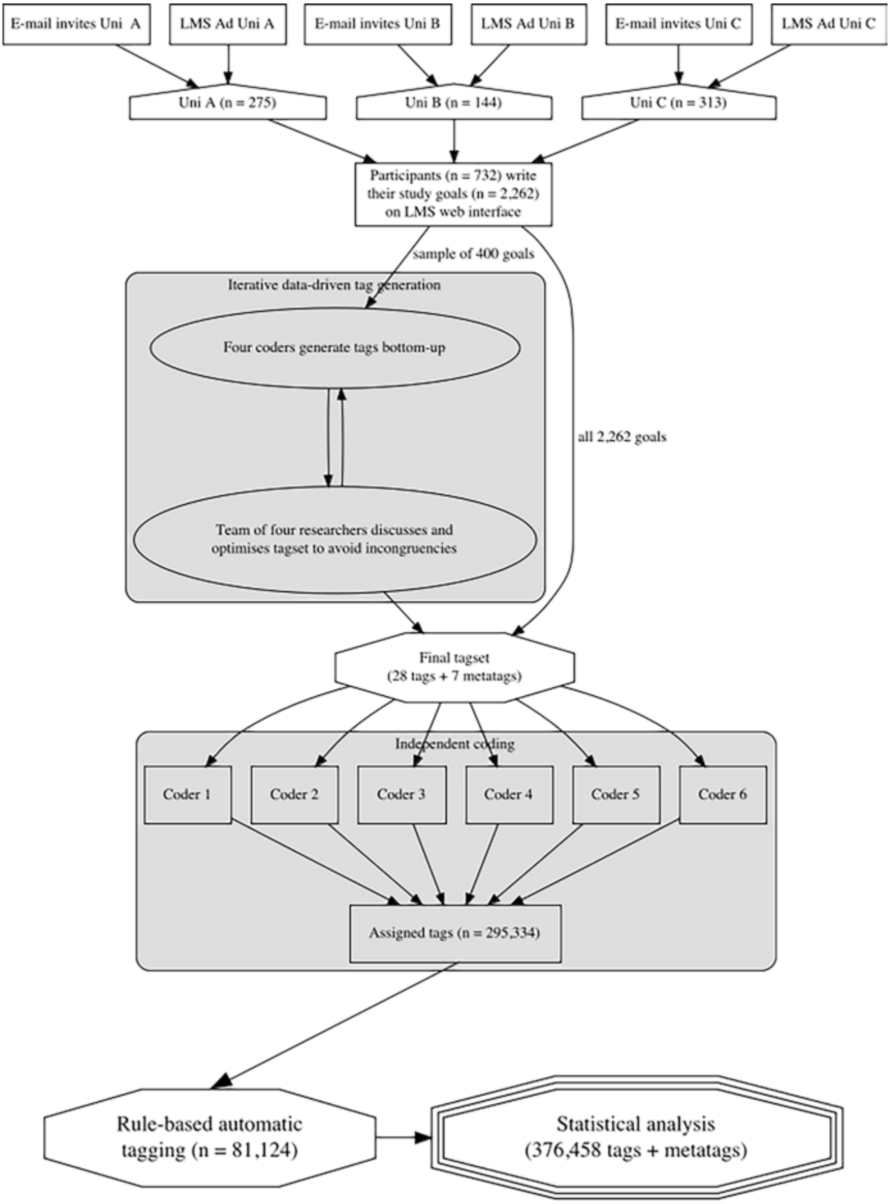


Fig. 4.1 The data processing pipeline as described in Weber and Le Foll (2020)

The second text box required participants to input their goals one by one. They also had the option of deleting previously submitted goals.

Post-data collection, a random sample of 400 goals were iteratively tagged by 1 of the researchers following a cyclical, data-driven process, and 4 project members discussed and improved the proposed tags. Thus, rather than imposing preconceived categories, commonalities and key themes were identified from the data (Braun & Clarke, 2006). Six independent raters then applied the final tagset to the full set of goals collected. They were instructed to make a binary decision for each goal/tag combination.

Post-manual coding, meta-tags were assigned rule-based and automated by a python script. For instance, if a rater assigned the tags ORIENTATION to the goal *Promotionsstelle bekommen* [get a PhD position], the meta-tag CAREER was automatically added. Additionally, following Bloom's (1956) taxonomy of educational objectives, the tags subsumed under the meta-tag *Education goals* were hypothesized to be sequentially dependent: PERSONAL GROWTH > COMPETENCES > COMPREHENSION > KNOWLEDGE. Consequently, if a goal was tagged as COMPETENCES, the tags KNOWLEDGE and COMPREHENSION were automatically added to the list of tags assigned to that goal.

Krippendorff's α (Krippendorff, 2004) was used to measure inter-rater agreement. This measure was used to quantify inter-rater agreement since it can handle missing data and can be compared to a number of other well-known metrics (Krippendorff, 2004). Relative tag frequencies were calculated for each tag across all ratings by all raters. Data analyses were carried out in Python 3.7 using the Pandas (McKinney, 2010), NumPy (Oliphant, 2006), and Statsmodels (Seabold & Perktold, 2010) libraries.

4.2.2 Technical Setup

The software used for data acquisition, storage, processing, and analysis was implemented by a software developer team of two research fellows and three student assistants as a web application, integrated into the local LMS.

The user interface for data collection (also referred to as frontend) was integrated into the learning management system *Stud.IP* (Stockmann & Berg, 2005) as a plugin coded in the PHP scripting language (see Kunda & Siame, 2017). The job of the plugin was exclusively to react to user input by sending data to the backend and to visualize the stored data.

The backend server runs on a virtual machine with the Ubuntu Linux operating system, as an application written in the *python3* programming language and based on the Django web framework (Django Software Foundation, 2013; Holovaty & Kaplan-Moss, 2007) and using a PostgreSQL database (Stonebraker & Kemnitz, 1991).

Backend and frontend communicate via a RESTful interface, using textual representations with a stateless protocol. As such, the backend is able to *receive*, *transfer*, *update*, or *delete* data through *requests* formatted, following the JSON-API standard. To ensure data security, data is transferred over SSL encrypted channels.

The goal tagging was implemented by HTML views provided by the backend. Each rater used personalized credentials to log in at any place and any time via web interface. Technically this setup was realized using in-built mechanisms of the Django framework, such as the Django template engine, the object-relational mapping, and Django's native authentication system.

4.2.3 Development of the Tagset

The iterative development process of the tagset was based on a preliminary data set of 400 goals and background knowledge about previous research. The final tagset consists of 28 tags, organized in 7 groups, each resulting in a meta-tag, which applies for a goal if at least 1 of the subsumed tags is assigned.

4.2.3.1 Private or Professional

The distinction between PRIVATE and PROFESSIONAL goals originates in our interest to know whether students pursue educational goals as a means to find a job and earn money or for personal reasons, such as thirst for knowledge or personal growth. Conceptually the distinction is related to the distinction between extrinsic and intrinsic motivation and also to the distinction between the two sets of CAREER GOALS (which can be considered as professional, because they are related to external rewards) and EDUCATIONAL GOALS (which can be considered as private because they are related to personal development).

4.2.3.2 Career Goals

The set of career goals contains specific formal achievements in an educational career as well as some more abstract kinds of goals related to professional success or earning a living in the future. As such, they are semantically related to performance goals in terms of the achievement goal literature, because they aim at benchmarks of the educational and societal surrounding[2].

GRADES. This tag is applied to goals that aim at a specific grade and also passing an exam or a course, as passing usually is equivalent to achieving at least a certain degree.

DURATION OF STUDIES. This tag is applied to goals which aim at completing a study program during a specific time point or within a certain time frame or generally as soon as possible.

[2] Except the *Orientation* tag, which can be considered to be an educational goal directed toward knowledge, comprehension, skills, and personal development in the domain of finding a path through a professional career

GRADUATION. This tag applied to a certain degree or professional position, such as for instance teacher or lawyer.

ORIENTATION. This tag applies to goals which aim at a clarification of one's degree, occupation, or other professional goals.

CAREER OPPORTUNITIES. This tag applies to all goals which aim to improve.

NETWORKING. This tag applies to goals that aim at establishing and strengthening personal contacts for the purpose of career building.

STATUS AND WEALTH. This tag aims at degrees, income, prestige, titles, power, resources, or possession.

SAFETY. This tag applies to goals which are directed to establishing material or professional safety.

4.2.3.3 Educational Goals

Educational goals are conceptually related to mastery goals in the achievement goal literature, insofar that learning or acquiring a competence or growing personally the essence of the goal.

The set of educational goal tags is inspired by Bloom's taxonomy of educational goals (Bloom et al., 1956), which is an early approach to structure educational objectives. The proposed taxonomy entails the six categories knowledge, comprehension, application, analysis, synthesis, and evaluation. From this set we adapted knowledge and comprehension and subsumed application, analysis, synthesis, and evaluation under the tag COMPETENCES because we understand all four as types of skills.

For the set of educational goals, we define a recursive subsumption relationship: Personal growth requires the acquisition of competences, which require at least a certain degree of comprehension which requires a certain degree of knowledge. In practice raters were instructed to assign only the highest goal tag in the subsumption hierarchy to avoid redundant mouse clicks. The subsumed lower tags were automatically complemented by a script post-tagging (see Fig. 4.2).

KNOWLEDGE. This tag applies to goals that are related to the acquisition of knowledge.

COMPREHENSION. This tag applies to goals which are related to understanding, which goes beyond knowledge and subsumes it.

COMPETENCES. This tag applies to goals that are related to the acquisition of an ability or a skill. This goes beyond comprehension and subsumes it.

PERSONAL GROWTH. Personal growth goals are directed toward self-improvement. Self-improvement motivation has been found to be a relevant motivational tendency in the domain of meta-cognition (Jiang & Kleitman, 2015). Growth goals have been shown to predict subjective well-being (SWB) in the future (Bauer & McAdams, 2010).

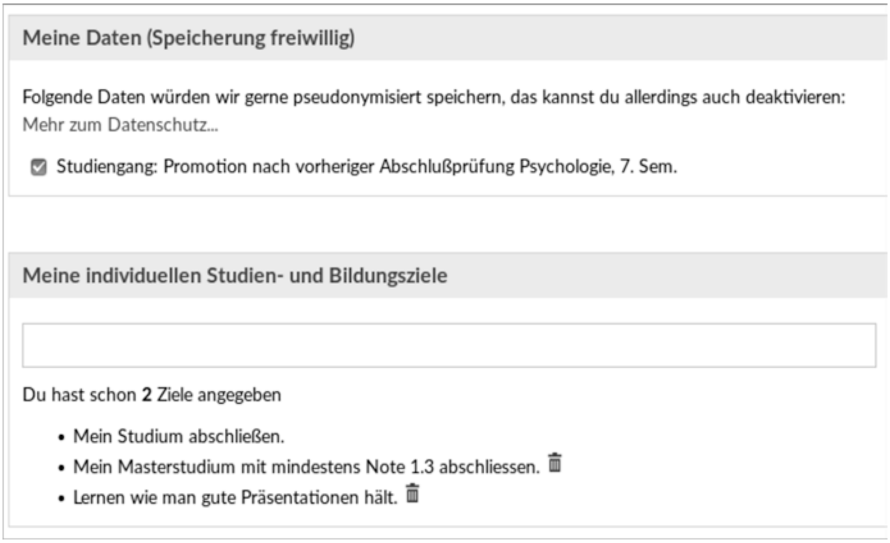


Fig. 4.2 The user interface offers a checkbox asking for user consent for study information data processing and a text input asking for goals in natural language. Inserted goals are displayed and can be deleted by clicking on the trash icon

4.2.3.4 Social Goals

Institutionalized learning in schools and the higher education system usually takes place in a group context of classes, courses, study programs, study groups, or homework groups. It has been demonstrated that perceived social support can positively affect motivation and achievement for educational goals (Song et al., 2015) It has been shown that public commitment to a goal improves goal commitment (Locke & Latham, 2002). Maslow's theory about basic human needs assumes a strong need for love and belongingness (Maslow, 1943), and positive psychology assumes that attachment and relatedness are central antecedents for eudaimonic well-being (see Ryan and Deci, 2001, for a review). In the set of goals, we identified two types of goals related to these needs:

COMMUNICATION AND CONTACT. This tag applies to goals that aim at meeting and communicating with other students or lecturers.

VOLUNTEER WORK AND IDEALISM. This tag applies to goals which are related to serving or working for a good cause. Examples are "working for an NGO" or "Getting engaged in the student's union."

4.2.3.5 Concrete Goals

The reason for this group of tags is the development of a digital study assistant for higher education. For each of the concrete goals, the digital study assistant is intended to derive recommendations and reminders. Therefore, we plan to train a machine learning model to detect those specific goals with data from this study.

PRACTICAL EXPERIENCES. This tag applies to internships and student jobs.

GOING ABROAD. This tag applies to students' interest in going abroad to study or do an internship. Universities provide a broad range of support services for such endeavors, which can be recommended by digital assistant.

FOREIGN LANGUAGES. This tag applies to language learning goals. On the basis of such a goal, the digital assistant can recommend language courses and exchange programs.

ACADEMIC AND SCIENTIFIC SKILLS. This tag applies to goals related to scientific methods, for which a broad range of extracurricular activities and learning opportunities exist, which are not obvious for students. A digital assistant can potentially unlock such learning opportunities.

PROGRAMMING SKILLS. This tag applies to goals aiming at programming or other computer-related skills. Digital technologies are an essential tool for all academic disciplines, but finding the right learning opportunities within the university may be a nontrivial task for students from nontechnical fields.

4.2.3.6 Temporal Horizon

The temporal scope of goals is a characteristic that has been investigated for a long time by goal-setting researchers. In the classical goal-setting literature by Locke and Latham (Latham & Brown, 2006; Latham & Locke, 1990; Locke & Latham, 2019), a dichotomous distinction between proximal and distal goals has been made.

The SIDDATA project aims at supporting students in the pursuit of educational goals by guiding processes of self-regulation on the micro-, meso-, and macro-level (Thelen et al., 2018). These terms refer to single learning sessions, learning organization, and time management during a semester and planning an education pathway through a study program or even lifelong learning.

In the context of an education goal-setting intervention, it is unlikely that students will formulate goals on the micro-level of just a few hours. The reason is that humans tend to commit only to goals that they assume to be within their abilities (Locke & Latham, 1990). So if a goal on the micro-level is feasible, it can be selected to be executed. If it is not feasible, then acquiring the required competences exceeds the micro time level, and the temporal scope of the resulting learning goal is higher. Consequently, time frames below one semester are not covered by the tagset. The following tags were used to operationalize the temporal horizon of educational goals:

WITHIN THIS SEMESTER: This tag reflects the educational environment which offers educational activities structured in semesters. In most cases these activities have a component of performance measure or examination which results in a certificate as a building block for a degree. We expected individual goals without curricular connection to be rare. So the rationale behind this tag is to apply it to courses and exams.

DURING STUDIES: This tag implies that the studies will be completed, which makes it applicable to learning goals without primarily curricular origin but individual interest. For example, courses that are perceived as especially challenging but mandatory for graduation are typical candidates for this tag.

POST-GRADUATION*:* This tag is applicable either to abstract distal learning goals, such as "understanding data science," or career goals with a scope beyond a study program, such as "becoming a data scientist."

4.2.3.7 Other Tags

This set of goal tags contains tags which do not fit into the other meta-tags, which are semantically related:

FUN, HAPPINESS AND SATISFACTION. This tag applies to goals with a hedonistic orientation. An example is "enjoying college life."

S.M.A.R.T. GOALS. Specific, measurable, assignable, realistic, and time-related goals were by Doran (Doran, 1981) in the context of management objectives to increase the clarity of goals and make them actionable and controllable. These aspects of goal setting obviously are important for estimations about the duration, effort, and circumstances of goal achievement. Goal specificity has been shown to increase the probability of goal achievement (Locke et al., 1989; Seijts et al., 2004).

TOO VAGUE. This tag is intended for goals which are so unspecific, that hardly any tag can be assigned. An example from the data set is "make experiences." Although one can subjectively imagine what might be meant by this formulation, in fact it is hard to be alive without making experiences. So this example is *too vague* to apply any tag, than this one.

NON-SENSICAL OR NON-GENUINE. This tag applies to goals that the raters expect not to be a serious goal but a joke. The goal "Gucci Socken" [Gucci socks] pointed at the necessity for such a tag. In some cases, the decision from an external perspective is nontrivial. For example, to the goal "Massenvernichtungswaffen" [weapons of mass destruction], the tag cannot clearly be assigned without knowing the context. Nevertheless, all raters tagged it a NON-SENSICAL, following their subjective view.

4.3 Results

In total, 732 students participated in the study. Among the participants, 74.69% agreed to provide data about their subject, degree type, and semester. 2262 goals were generated so that, on average, participants provided around 3 goals each. The length of goals varied from single words to several elaborate sentences with word counts from 1 to 39 and an average of 3 words. The number of characters per goal ranged from 3 to 276 with an average of 27. In total, 2262 goals were tagged. Due to time and resource constraints, the six raters did not all tag the full data set. A total of 295.334 manual ratings were made. Post-coding, the data cleaning process involved excluding goals identified by the raters as NON-SENSICAL utterances and NON-GENUINE goals ($\alpha = 0.712$), e.g., *Gucci Socken* [Gucci socks], and resulted in a total of 2.204 goals to be further analyzed. The results are summarized in Figs. 4.4 and 4.5, which show relative frequencies in percent and Krippendorff's α as measure for inter-rater agreement.

4.3.1 Inter-rater Agreement

The overall very low inter-rater agreement rates reveal that, in practice, many of the tags proved rather difficult to distinguish. Thus, it was originally assumed, in line with Ahn et al.'s (2012) hierarchical goal model, that every goal would be assigned the meta-tag PRIVATE OR PROFESSIONAL; in other words, that every goal would be classified as either related to (future) private or professional life plans. In practice, however, 57.19% of goals were not assigned this meta-tag. By contrast, there even were some cases where both the PRIVATE and the PROFESSIONAL tags were assigned to the same goal. Thus, it would appear that educational goals are often tied to both personal and professional interests and that, in many cases, the two are not easily disentangled.

Conceived as a pragmatic way to make the fuzzy concept of distal and proximal goals operational, the temporal scope of goals was also often hard to determine without additional background information. The α scores for the three temporal tags WITHIN THIS SEMESTER, DURING STUDIES, and POST-GRADUATION are among the lowest. Theoretically, these three tags cover all the possible temporal scopes of study goals. Hence, in principle, at least one of the tags should apply to each goal formulated. However, as many as 67.14% of goals were not assigned a temporal tag. This was due, on the one hand, to participants not assigning clear temporal scopes to their goals and, on the other hand, to implicit temporal scopes not being inferred by the raters.

Similarly, the inter-rater agreement rates reported for the EDUCATIONAL GOALS tags are also surprisingly low. This is largely due to participants' often very sparse elaboration of goals. Thus, many participants formulated goals that simply read *Lernen [learn/revise]* or *Neue Inhalte erlernen [learn new things]*, for which even

PRIVATE OR PROFESSIONAL	CAREER GOALS	EDUCATIONAL GOALS	SOCIAL GOALS	CONCRETE GOALS	TEMPORAL HORIZON	OTHER TAGS
PROFESSIONAL	GRADES	KNOWLEDGE	COMMUNICATION AND CONTACT	WORK(-RELATED) EXPERIECES	WITHIN THIS SEMESTER	FUN, HAPPINESS AND SATISFACTION
PRIVATE	DURATION OF STUDIES	COMPREHENSION	VOLUNTEER WORK AND IDEALISM	GOING ABROAD	DURING STUDIES	S.M.A.R.T. GOALS
	GRADUATION	COMPETENCES		FOREIGN LANGUAGES	POST-GRADUATION	TOO VAGUE
	ORIENTATION	PERSONAL GROWTH		ACADEMIC AND SCIENTIFIC SKILLS		NON-SENSICAL OR NON-GENUINE
	CAREER OPPORTUNITIES			PROGRAMMING SKILLS		
	NETWORKING					
	STATUS AND WEALTH					
	SECURITY					

Fig. 4.3 The 28 tags are grouped into 7 groups, each forming a meta-tag

the four broad Educational Goals tags of the present tagset (Fig. 4.3) were already too detailed.

The Safety (referring to job and financial safety) tag was also frequently difficult to ascertain. Again, goals that explicitly mention these factors are rare, e.g., *Ein eigenes Haus haben [be a homeowner].*

Raters also reported that the tags Networking (assigned to 46 goals) and Social contact (assigned to 76 goals) were hard to distinguish from one another. The tagset included this distinction in the hope of tapping into the motives behind the two goal types. On a pragmatic level, however, raters frequently lacked contextual information to disambiguate the 2 so that 22 goals were assigned both the Networking and Social contact tags. It is worth noting that, in German culture, admitting to building relationships purely for professional reasons is often not regarded as socially acceptable and may be perceived as selfish.

On the other hand, Concrete Goals tags, such as those referring to learning a foreign language, studying abroad, graduating quickly, acquiring programming skills, obtaining good grades, and gathering work experience, have a high inter-rater agreement. This is probably due to their specificity and the fact that they are very frequent and, consequently, familiar to the raters, who were university students themselves.

4.3.2 Relative Frequencies

Due to the number of students (n = 732) and, in particular, the uncontrolled variables in the selection of the participants, the external validity of the relative frequencies presented in Fig. 4.5 toward a generalization to the goal characteristics of university students is inherently limited. Nevertheless, they reveal that the most frequent tags assigned to participants' self-set goals are Career, Knowledge, During Studies, Competences, and Graduation. Career, here, refers to goals clearly related to either studies or work, as opposed to Private goals; hence this

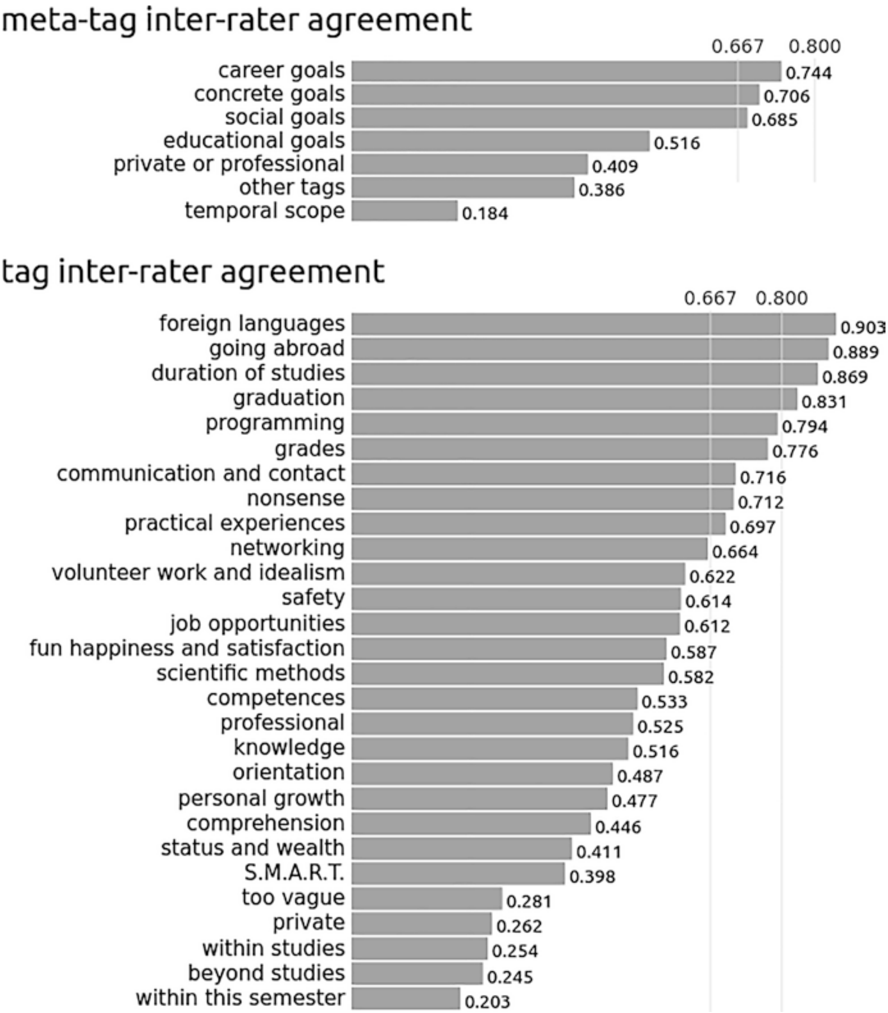


Fig. 4.4 Krippendorff's α as measure for inter-rater agreement of all tags and meta-tags (interpretation: $\alpha \geq 0.800$ tentative conclusive, $\alpha \geq 0.667$ acceptable according to Krippendorff, 2004)

finding was to be expected. The fact that many of the goals were also tagged as DURING STUDIES suggests that the majority of students' goals do refer to their current student status, as opposed to more long-term goals referring to their professional lives beyond their studies.

Indeed, many of the participants' goals revolved around learning objectives and were thus assigned the tag KNOWLEDGE. However, these goals were often highly unspecified: many simply stating *Wissen [knowledge]* or *Lernen [learning]*, courses, or broad disciplines, e.g., *Biologie [biology]* and *Statistik bestehen [pass statistics]*. Other students formulated longer, but even more general goals such as *mehr Wissen*

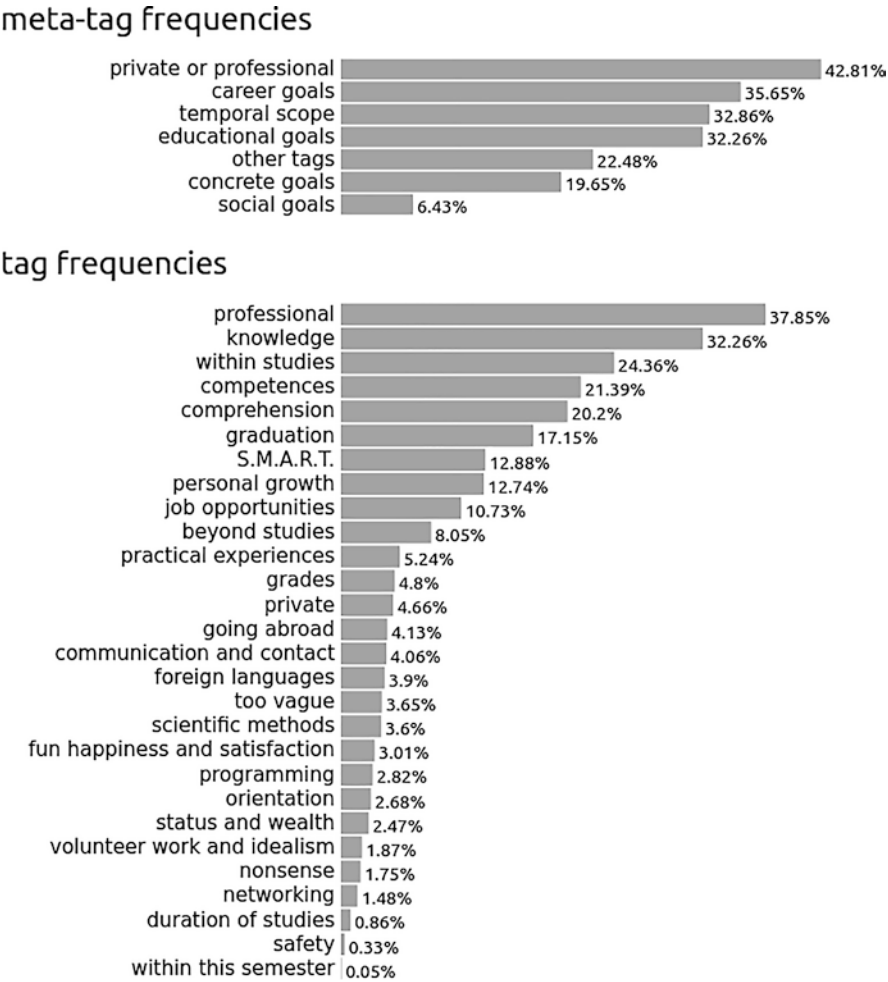


Fig. 4.5 Relative frequencies for tags and meta-tags

sammeln, allgemein und fachspezifisch [gather more general and subject-specific knowledge].

Nevertheless, a relatively high proportion of goals were also assigned the tag COMPETENCES, which refers to the application of comprehended knowledge. Such goals were often more specific, so that the tag COMPETENCES is highly correlated with the tags corresponding to specific skills such as PROGRAMMING and FOREIGN LANGUAGES skills, though it was also frequently assigned to goals referring to critical thinking and soft skills, for which no specific tags are included in the present tagset. Across all degree programs, many students articulated a wish to improve their foreign language skills. It is striking that such goals are also frequently formulated in a very abstract manner, e.g., *eine weitere Sprache erlernen [to learn another language]*.

Given these examples, it will come as no surprise that the relative frequency of SMART goals is low. The inter-rater agreement rate for SMART goals is also surprisingly low: this is due to a disagreement between the raters as to whether goals referring to Graduation, which were also among the most frequent, e.g., *Master abschließen* [*complete my Masters*], should be considered specific, measurable, and time-bound.

4.4 Discussion

This study encouraged students to submit their personal study goals via an online interface embedded in universities' local LMS. Previous research has shown that the very process of formulating such goals may induce learning and contribute to actually reaching these goals (e.g., Locke and Latham, 2002; Schippers et al., 2020). In particular, Morisano et al.'s (2010) study concluded that detailing personal goals and strategies can significantly improve educational performance. Thus, such personal goal-setting interventions can both contribute to making the value of goals more salient and help develop strategies to attain them.

We suggest that a digital study assistant could provide this kind of support, at large scale, in a personalized manner. Further, such a tool could support the monitoring of goal progress, which is also known to improve effective self-regulation and increase the likelihood of successful goal outcomes (Harkin et al., 2016). The present results suggest that Concrete goals may be most easily supported by a digital study assistant on the basis of simple rule-based algorithms. The tagged data from our study can serve as labeled training data for machine learning algorithms capable to assign tags to goals entered as user input.

First steps toward the implementation of the long-term goal of tagging students' educational goals automatically in a digital data-driven study assistant software have been taken. We used the labeled goal data from the study summarized in this chapter, to train and evaluate a pre-trained Bidirectional Encoder Representations from Transformers (BERT) (Devlin et al., 2019) machine learning model. Preliminary results show that the performance of the artificial intelligence agent equals the performance of a swarm of agents with natural intelligence insofar, that the accuracy highly differs between tags. This finding is in line with methodological considerations concerning external characterization of goal characteristics: Some goal characteristics are goal-inherent and suitable for detection from an external perspective; some have subjective qualities which can only be measured by self-assessment and even may be unstable over time.

4.4.1 Methodological Considerations

An implicit assumption we made in the design of the current study is goals can be objectively characterized by external raters. From a retrospecive perspective, this assumption is in a clear contradiction to the—by definition—subjective nature of goals. The same goal formulation may have different meanings for different persons because personal preferences and predispositions determine why a goal is worth striving for. Taking math classes may be a pleasurable experience for some persons and an unpleasant means toward a superordinate goal for others. So in this specific example, only the person having the goal can provide reliable information about the motivational background and properties of a goal. Raters, on the other hand, lack essential information and may tend to compensate this by projecting their preferences and predispositions onto the goal.

The method of assigning goal characteristics as tags externally seems to work well for very specific types of goals, such as GOING ABROAD, GRADUATION, or LEARNING A FOREIGN LANGUAGE, but not so well for more abstract characteristics, such as the TEMPORAL SCOPE of a goal or whether it is PRIVATE or PROFESSIONAL.

4.4.2 Outlook

This study has made clear that, without any support, students tend to formulate very brief, unspecified goals, whereas meta-analyses have shown that specific and challenging, yet attainable, goals are most likely to be reached (Locke and Latham, 2002). Alternatively, Schippers et al. (2020) hypothesize that developing detailed strategies for goal attainment may compensate for a lack of specificity in the goal formulated. In either cases, the results of our study suggest that university students could benefit from additional support to both formulate their goals and develop specific strategies to attain them.

An approach that we think might work well in the context of a digital study assistant for higher education is hierarchical goal systems: Originating from a personally relevant distal educational goal, sub-goals are derived until the action level is reached. The procedure is similar to the personal projects analysis (PPA) procedure of laddering, where participants are repeatedly asked for sub-goals of superordinate goals (Little, 1983). There are good arguments to assume that hierarchical structures underlie the control of sequential human behavior (Cooper & Shallice, 2006).

As mentioned in Sect. 4.1, the overall inter-rater agreement on the externally assigned tags is relatively low. This finding raises methodological questions concerning external characterization of goals, especially for highly subjective aspects. Goal characteristics which cannot be assessed from an external perspective merely from an articulated goal itself can better be captured by psychometric self-report measures used by the person with the goal in mind. The ongoing development of a goal characteristics questionnaire (Iwama et al., 2019), which constitutes a psychometric measure covering a broad variety of goal-related variables, bears a huge potential for future research on goals.

Acknowledgments The authors would like to express their deep appreciation to the raters who spent hours with assigning tags to more or less entertaining goals. We would also like to thank Elen Le Foll for her support, especially with Fig. 4.1 which elucidates the data processing procedure.

References

Ahn, H., Clermont, M., Dyckhoff, H., & Höfer-Diehl, Y. (2012). Entscheidungsanalytische Strukturierung fundamentaler Studienziele. *Zeitschrift für Betriebswirtschaft, 82*(11), 1229–1257. https://doi.org/10.1007/s11573-012-0625-5

Alessandri, G., Borgogni, L., Latham, G. P., Cepale, G., Theodorou, A., & De Longis, E. (2020). Self-set goals improve academic performance through nonlinear effects on daily study performance. *Learning and Individual Differences, 77*, 101784. https://doi.org/10.1016/j.lindif.2019.101784

Amer, A. (2006). Reflections on bloom's revised taxonomy. *Electronic Journal of Research in Educational Psychology, 4*(8), 213–230.

Ames, C. (1992). Classrooms: Goals, structures, and student motivation. *Journal of Educational Psychology, 84*(3), 261–271. https://doi.org/10.1037/0022-0663.84.3.261

Ames, C., & Archer, J. (1987). Mothers' beliefs about the role of ability and effort in school learning. *Journal of Educational Psychology, 79*(4), 409–414. https://doi.org/10.1037/0022-0663.79.4.409

Ballard, T., Yeo, G., Loft, S., Vancouver, J. B., & Neal, A. (2016). An integrative formal model of motivation and decision making: The MGPM*. *Journal of Applied Psychology, 101*(9), 1240–1265. https://doi.org/10.1037/apl0000121

Bauer, J. J., & McAdams, D. P. (2010). Eudaimonic growth: Narrative growth goals predict increases in ego development and subjective well-being 3 years later. *Developmental Psychology, 46*(4), 761–772. https://doi.org/10.1037/a0019654

Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). Taxonomy of educational objectives. In B. S. Bloom (Ed.), *The classification of educational goals*. David McKay. https://www.uky.edu/~rsand1/china2018/texts/Bloom%20et%20al%20-Taxonomy%20of%20Educational%20Objectives.pdf

Boniwell, I., & Zimbardo, P. G. (2012). Balancing time perspective in pursuit of optimal functioning. In P. A. Linley & S. Joseph (Eds.), *Positive psychology in practice* (pp. 165–178). Wiley. https://doi.org/10.1002/9780470939338.ch10

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Cooper, R. P., & Shallice, T. (2006). Hierarchical schemas and goals in the control of sequential behavior. *Psychological Review, 113*(4), 887–916. https://doi.org/10.1037/0033-295X.113.4.887

Dallmayr, F. (2008). Desire and the desirable: Nicholas de Cusa's interfaith imagination. *Religion and the Arts, 12*, 1–3. https://doi.org/10.1163/156852908X271178

Deci, E. L., & Ryan, R. M. (1985). Intrinsic motivation and self-determination in human behavior. In *Intrinsic motivation and self-determination in human behavior*. Springer. https://doi.org/10.1007/978-1-4899-2271-7

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In *NAACL HLT 2019—2019 Conference of the*

North American Chapter of the Association for Computational Linguistics: Human Language Technologies—Proceedings of the Conference, *1*.

Django Software Foundation. (2013). Django: The Web framework for perfectionists with deadlines. Djangoproject.Com.

Doran, G. T. (1981). There's a S.M.A.R.T way to write management's goals and objectives. *Management Review, 70*(11), 35–36.

Dweck, C. S. (1986). Motivational processes affecting learning. *The American Psychologist*, 1040–1048. https://doi.org/10.1037/0003-066X.41.10.1040

Ehrlich, C. (2012). Be careful what you wish for but also why you wish for it—Goal-striving reasons and subjective well-being. *Journal of Positive Psychology, 7*(6), 493–503. https://doi.org/10.1080/17439760.2012.721382

Elliot, A. J. (1999). Approach and avoidance motivation and achievement goals. *Educational Psychologist, 34*(3), 169–189. https://doi.org/10.1207/s15326985ep3403_3

Elliot, A. J., & Church, M. A. (1997). A hierarchical model of approach and avoidance achievement motivation. *Journal of Personality and Social Psychology, 72*(1), 218–232. https://doi.org/10.1037/0022-3514.72.1.218

Elliot, A. J., & Harackiewicz, J. M. (1996). Approach and avoidance achievement goals and intrinsic motivation: A mediational analysis. *Journal of Personality and Social Psychology, 70*(3), 461–475. https://doi.org/10.1037/0022-3514.70.3.461

Elliot, A. J., & McGregor, H. A. (2001). A 2 × 2 goal achievement framework. *Journal of Personality and Social Psychology, 80*(3), 501–519.

Elliot, A. J., & Thrash, T. M. (2001). Achievement goals and the hierarchical model of achievement motivation. *Educational Psychology Review, 13*(2), 139–156.

Elliot, A. J., Murayama, K., & Pekrun, R. (2011). A 3 × 2 achievement goal model. *Journal of Educational Psychology, 103*(3), 632–648. https://doi.org/10.1037/a0023952

Freund, A. M., Hennecke, M., & Mustafić, M. (2012). On gains and losses, means and ends: Goal orientation and goal focus across adulthood. In R. M. Ryan (Ed.), *The Oxford handbook of human motivation*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780195399820.013.0016

Harkin, B., Webb, T. L., Chang, B. P. I., Prestwich, A., Conner, M., Kellar, I., Benn, Y., & Sheeran, P. (2016). Does monitoring goal progress promote goal attainment? A meta-analysis of the experimental evidence. *Psychological Bulletin, 142*(2), 198–229. https://doi.org/10.1037/bul0000025

Heckhausen, J., Wrosch, C., & Schulz, R. (2010). A motivational theory of life-span development. *Psychological Review, 117*(1), 32. https://doi.org/10.1037/a0017668

Holovaty, A., & Kaplan-Moss, J. (2007). The Django book. *Interface*.

Iwama, G., Wirzberger, M., & Lieder, F. (2019). *The Goal Characteristics (GC) questionnaire: A comprehensive measure for goals' content, attainability, interestingness, and usefulness*. Max Planck Institute for Intelligent Systems. https://doi.org/10.13140/RG.2.2.24237.84961

Jiang, Y., & Kleitman, S. (2015). Metacognition and motivation: Links between confidence, self-protection and self-enhancement. *Learning and Individual Differences, 37*, 222–230. https://doi.org/10.1016/j.lindif.2014.11.025

Krippendorff, K. (2004). *Content analysis: An introduction to its methodology* (2nd ed.). SAGE.

Kruglanski, A. W., Shah, J. Y., Fishbach, A., Friedman, R., Woo Young Chun, & Sleeth-Keppler, D. (2002). A theory of goal systems. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. *34*, pp. 331–378). Academic Press. doi:https://doi.org/10.1016/s0065-2601(02)80008-9

Kunda, D., & Siame, A. (2017). Evolution of PHP applications: A systematic literature review. *International Journal of Recent Contributions from Engineering, Science & IT (IJES), 5*(1), 28–39.

Kung, F. Y. H., & Scholer, A. A. (2018). A network model of goals boosts convergent creativity performance. *Frontiers in Psychology, 9*(OCT), 1–12. https://doi.org/10.3389/fpsyg.2018.01910

Latham, G., & Brown, T. (2006). The effect of learning, distal, and proximal goals on MBA self-efficacy and satisfaction. *Applied Psychology: An International Review, 55*(4), 6060–6123. http://web.ebscohost.com/ehost/detail?vid=36&hid=117&sid=827b8f48-1448-4d4a-9bd2-0866d9194f0b%40sessionmgr104&bdata=JnNpdGU9ZWhvc3QtbGl2ZQ%3D%3D#db=psyh&AN=2006-20408-006

Latham, G. P., & Locke, E. A. (2007). New developments in and directions for goal-setting research. *European Psychologist, 12*(4), 290–300. https://doi.org/10.1027/1016-9040.12.4.290

Little, B. R. (1983). Personal projects: A rationale and method for investigation. *Environment and Behavior, 15*(3), 273–309. https://doi.org/10.1177/0013916583153002

Locke, E. A., & Latham, G. P. (1990). *A theory of goal setting and task performance*. Prentice Hall.

Locke, E. A., & Latham, G. P. (2002). Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. *American Psychologist, 57*(9), 705–717. https://doi.org/10.1037/0003-066X.57.9.705

Locke, E. A., & Latham, G. P. (2019). The development of goal setting theory: A half century retrospective. *Motivation Science, 5*(2). https://doi.org/10.1037/mot0000127

Locke, E. A., Chah, D. O., Harrison, S., & Lustgarten, N. (1989). Separating the effects of goal specificity from goal level. *Organizational Behavior and Human Decision Processes, 43*(2), 270–287. https://doi.org/10.1016/0749-5978(89)90053-8

Maslow, A. H. (1943). A theory of human motivation. *Psychological Review, 50*(4), 370–396. https://doi.org/10.1037/h0054346

McKinney, W. (2010). Data structures for statistical computing in Python. *Proceedings of the 9th Python in Science Conference*, pp.51–56. http://conference.scipy.org/proceedings/scipy2010/mckinney.html

Morisano, D. (2008). *Personal goal setting in university students: Effects on academic achievement, executive functioning, personality, and mood*. McGill University.

Morisano, D. (2013). Goal setting in the academic arena. In E. A. Locke & G. P. Latham (Eds.), *New developments in goal setting and task performance* (pp. 495–506). Routledge/Taylor & Francis Group. http://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=2013-00428-030&site=ehost-live

Morisano, D., Hirsh, J. B., Peterson, J. B., Pihl, R. O., & Shore, B. M. (2010). Setting, elaborating, and reflecting on personal goals improves academic performance. *Journal of Applied Psychology, 95*(2), 255–264. https://doi.org/10.1037/a0018478

Murayama, K., Elliot, A. J., & Friedman, R. (2012). Achievement goals. In R. M. Ryan (Ed.), *The Oxford handbook of human motivation* (pp. 190–207). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780195399820.013.0012

Murphy, P. K., & Alexander, P. A. (2000). A motivated exploration of motivation terminology. *Contemporary Educational Psychology, 25*(1), 3–53. https://doi.org/10.1006/ceps.1999.1019

Nicholls, J. G. (1984). Achievement motivation: Conceptions of ability, subjective experience, task choice, and performance. *Psychological Review, 91*(3), 328–346. https://doi.org/10.1037/0033-295X.91.3.328

Nicholls, J. G. (2017). Conceptions of ability and achievement motivation: A theory and its implications for education. In S. G. Paris, G. M. Olson, & H. W. Stevenson (Eds.), *Learning and motivation in the classroom.* Routledge. https://doi.org/10.4324/9781315188522

Oates, B. J. (2006). *Researching information systems and computing*. SAGE.

Oliphant, T. E. (2006). *A guide to NumPy*. Trelgol. http://web.mit.edu/dvp/Public/numpybook.pdf

Olos, L., Hoff, E.-H., & Härtwig, C. (Eds.). (2014). *Berufliche Zielklärung und Selbststeuerung*. Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-03650-8

Pintrich, P. R. (2000a). An achievement goal theory perspective on issues in motivation terminology, theory, and research. *Contemporary Educational Psychology, 25*(1), 92–104. https://doi.org/10.1006/ceps.1999.1017

Pintrich, P. R. (2000b). Multiple goals, multiple pathways: The role of goal orientation in learning and achievement. *Journal of Educational Psychology, 92*(3), 544–555. https://doi.org/10.1037/0022-0663.92.3.544

Ryan, R. M., & Deci, E. L. (2001). On happiness and human potentials: A review of research on hedonic and eudaimonic well-being. *Annual Review of Psychology, 52*(1), 141–166. https://doi.org/10.1146/annurev.psych.52.1.141

Schippers, M. C., Morisano, D., Locke, E. A., Scheepers, A. W. A., Latham, G. P., & de Jong, E. M. (2020). Writing about personal goals and plans regardless of goal type boosts academic performance. *Contemporary Educational Psychology, 60*(November 2019), 101823. https://doi.org/10.1016/j.cedpsych.2019.101823

Seabold, S., & Perktold, J. (2010). statsmodels: Econometric and statistical modeling with python. In *9th Python in Science Conference*.

Seijts, G. H., Latham, G. P., Tasa, K., & Latham, B. W. (2004). Goal setting and goal orientation: An integration of two different yet related literatures. *Academy of Management Journal, 47*(2), 227–239. https://doi.org/10.2307/20159574

Song, J., Bong, M., Lee, K., & Kim, S. (2015). Longitudinal investigation into the role of perceived social support. *Journal of Educational Psychology, 107*(3), 821–841.

Stockmann, R., & Berg, A. (2005). Stud.IP. *Eleed, 1*(1) https://www.studip.de/

Stonebraker, M., & Kemnitz, G. (1991). The POSTGRES next generation database management system. *Communications of the ACM, 34*(10), 78–92. https://doi.org/10.1145/125223.125262

Thelen, T., Osada, S., & Weber, F. (2018). *SIDDATA—Joint project for individualization of studies through digital, data-driven assistants*. Universität Osnabrück. https://www.siddata.de/en/

Travers, C. J., Morisano, D., & Locke, E. A. (2015). Self-reflection, growth goals, and academic outcomes: A qualitative study. *British Journal of Educational Psychology, 85*(2), 224–241. https://doi.org/10.1111/bjep.12059

Vancouver, J. B., & Austin, J. T. (1996). Goal constructs in psychology: Structure, process, and content. *Psychological Bulletin, 120*(3), 338–375.

Virtanen, P., Nevgi, A., & Niemi, H. (2013). Self-regulation in higher education: Students' motivational, regulational and learning strategies, and their relationships to study success. *Studies for the Learning Society, 3*(1–2), 20–34. https://doi.org/10.2478/sls-2013-0004

Weber, F. (2019). Goal trees as structuring element in a digital data-driven study assistant. In D. G. Sampson, D. Ifenthaler, I. Pedro, & M. L. Mascia (Eds.), *16th International Conference on Cognition and Exploratory Learning in Digital Age (CELDA 2019) (Issue Celda)* (pp. 413–416). IADIS Press. https://doi.org/10.33965/celda2019_201911c053

Weber, F., & Le Foll, E. (2020). A tagset for university students' educational goals. In D. G. Sampson, D. Ifenthaler, I. Pedro, & P. Isaías (Eds.), *17th International Conference on Cognition and Exploratory Learning in Digital Age (CELDA 2020) (Issue Celda)* (pp. 27–34). IADIS Press. https://doi.org/10.33965/celda2020_202014l004

Wellington, J. (2020). Bloom's taxonomy. In *Secondary education: The key concepts*. Routledge. doi:https://doi.org/10.4324/9780203488317-7

Zimmerman, B. J. (1989). Models of self-regulated learning and academic achievement. In B. J. Zimmerman & D. H. Schunk (Eds.), *Self-regulated learning and academic achievement: Theory, research and practice* (pp. 1–25). Springer. https://doi.org/10.1007/978-1-4612-3618-4_1

Zimmerman, B. J. (1990). Self-regulating academic learning and achievement: The emergence of a social cognitive perspective. *Educational Psychology Review, 2*(2), 173–201. http://www.jstor.org.libraryproxy.griffith.edu.au/stable/pdfplus/23359581.pdf

Chapter 5
A Digital Environment for University Guidance: An Analysis of the Academic Results of Students Who Practice Self-Assessment in Orient@mente, an Open Online Platform to Facilitate the Transition from Secondary School to Higher Education

Francesco Floris, Marina Marchisio, Sergio Rabellino, and Matteo Sacchet

5.1 Introduction

This chapter is an extension of work originally presented in the 17th International Conference on Cognition and Exploratory Learning in the Digital Age (CELDA) (Floris, Marchisio, Rabellino, & Sacchet, 2020). Massive open online courses (MOOCs) are a pervasive instrument, appearing now in many aspects of education. Keywords related to MOOCs are open education, open access, interactivity, and digital competencies. Education via MOOCs is where many pedagogical studies on user interaction with online systems were born. In addition to education, MOOCs can also pursue goals that are unrelated to disciplinary contents. As an example, with the help of an automatic assessment system, institutions can enhance students' evaluation and management of their self-regulated learning. With the help of computing environments, teachers can enhance problem-solving skills. With the help of web conference tools, peers can collaborate and enhance team-working skills across different regions and countries.

In the 2010 report, the European Commission "Europe 2020" (European Commission, 2010) provided a strategy for smart, sustainable, and inclusive growth of Europe. One of the objectives about education is the reduction of the early school

F. Floris · M. Marchisio · M. Sacchet (✉)
Department of Molecular Biotechnology and Health Sciences, University of Turin, Turin, Italy
e-mail: matteo.sacchet@unito.it

S. Rabellino
Department of Computer Science, University of Turin, Turin, Italy

D. Ifenthaler et al. (eds.), *Orchestration of Learning Environments in the Digital World*, Cognition and Exploratory Learning in the Digital Age,
https://doi.org/10.1007/978-3-030-90944-4_5

leaver rate (percentage of 18–24-year-olds in the EU who completed at most a lower secondary education and were not in further education or training) by 5 percentage points, from 15% in 2010 to a target of 10%. Another objective is the increase in the percentage of young graduates aged between 30 and 34 by approximately 10 percentage points, from 31% in 2010 up to 40%. In 2019, the percentage of early school leavers in Europe averages at 10.2%, with quite large differences across countries, ranging from 3.0% in Croatia to 17.3% in Spain (Early leavers from education and training, 2021). On the other side, in 2019, more than 40% of the 30–34-year-olds in the EU had completed tertiary education (Educational attainment statistics, 2021).

There are a close connection and many similarities between these rates with the difficulties that students encounter in the transition phase from secondary school to university, that is, the secondary to tertiary transition (STT). Students radically change their perspectives. They need to be aware of the choice that will affect their future and face the difference between expectations and reality, which can lead to unexpected failures of the learning strategies developed in previous education. The education systems in all countries must make this challenging experience as smooth as possible. Technologies and virtual environments can support students in this phase because students can have a preview of what they will face at university in advance. Online assessments containing feedback allow students to check whether their knowledge or their attitude is adequate to attend a specific university program, and interactive learning materials can help them fill their gaps.

The main mission of university orientation is the reduction of the dropout rate, which is a measure that directly affects the evaluation and the ranking of the university itself. Another mission of university orientation is the balance between the demand for professionals by the job market and the number of new graduates, to secure a successful transition into the world of work. Thus, many degree programs have restricted access: they require students to take an admission test in order to control the number of new students, which can enter the program according to the score they get. Other programs, without restrictions, assess just the minimum requirements, and, in case of failure, they deliver remedial courses to align all the students to a common threshold. These assessments can give students a response on their attitude about the discipline, but this is not their main purpose, which is instead a summative evaluation of the students to evaluate their eligibility. Moreover, universities promote various initiatives and different guidance activities for students. Orientation days, job fairs, and open days are the most common initiatives promoted by universities, education offices, and secondary schools. In addition, technologies offer a solution for guidance activities, providing a massive dissemination of information and automatic assessment inside digital learning environments (DLE), together with multimedia resources and interactive components.

The University of Turin proposes several strategies and actions to facilitate the STT using new technologies. The model consists of a DLE, integrated with an advanced computing environment (ACE), an automatic assessment system (AAS), and a web conference tool. The DLE delivers self-paced open online courses (OOCs). In our model there is no M for massive since the focus of these strategies is on the single student. Moreover, differently from most MOOCs, there is no restriction on when to access the courses: students can start attending anytime and anywhere, with any device with Internet connection. In addition to these advantages on availability, the basin of attraction is larger than other kinds of activities: considering the users of the various actions, more than 80,000 users were involved in 5 years of deployment, with possible duplicate across the platforms, but around 67,000 distinct students, from the most populated platform. On average, every month there are around 1700 unique logins, corresponding to more than 60,000 sign-ins per month.

Thus, the DLE collects many data, which are extremely useful to analyze and evaluate the impact of the orientation actions and improve and enhance their effect. In this paper, we considered the number of first-year university credits (ECTS) obtained by students at the University of Turin, comparing this number with subscription, access, and attendance of at least one course in the Orient@mente platform. We will analyze and discuss what emerges from data collected from different sources.

Section 5.2 describes the state of the art about university orientation. Section 5.3 states the research questions and explains the methodology adopted for the investigation, together with the description of actions performed on data. Section 5.4 shows the results and discusses the relevant implications.

5.2 State of the Art

Every time a student fails or drops out of school or university, the whole education system fails, and this can happen at any level. Thus, all institutions aim at preventing failures and enhancing student success. There are numerous strategies to reach this objective, such as addressing evaluation, grouping, motivation, practices, responsibility, and relationships. According to Krull and Duart (2018), interventions in student orientation split into two main branches: academic support, connected to the development of cognitive and learning skills, and nonacademic support, which focuses on emotional and organizational aspects such as counselling and guidance services. When performed correctly, both branches have a positive impact on the students' experience.

Data-driven decisions can guide the choice on which strategy most suits institutions. In a fast-changing world, descriptive analytics, predictive and prescriptive models suggest which changes or strategies are good for preventing failure and enhancing success. When institutions can detect possible failures in advance, interventions prevent students' failure. Prediction algorithms using data mining techniques are one of the latest trends (Marquez-Vera, Morales, & Soto, 2013), but also, more in general, learning analytics (LA) (Gašević, Dawson, & Siemens, 2015), which are now a trend topic in conferences (Ochoa & Merceron, 2018). In Schumacher and Ifenthaler (2018), findings emphasize the need to design highly personalized and adaptable learning analytics systems to meet students' needs. LA approaches make use of machine learning algorithms, too. This kind of algorithms usually works as black boxes: they provide an output without any clue or information on the decision process. Research in this topic needs further investigations, concerning the ethical aspects and fairness (Riazy, Simbeck, & Schreck, 2020). Developers should keep in mind that predictions of a student's performance directly affect the student's success. Researchers developed methods and tutorials to detect discrimination (Hajian, Bonchi, & Castillo, 2016), even though it is endemic: even with no discrimination intention in the developer, the algorithm has biases.

Some experiences join orientation and LA. In Lonn, Aguilar, and Teasley (2015), the authors address summer programs designed to improve retention and academic success among at-risk postsecondary students, and they investigate how a learning analytics-based intervention empowers academic advisors during their face-to-face meetings with students. The findings suggest that the design of learning analytics interventions should consider students' perceptions of their goals and formative performance, since they affect students' interpretations of their own data as well as their subsequent academic success.

More power comes from digital technologies and online learning, and their importance is stressed in the Digital Education Action Plan 2021–2027 (DEAP, 2020). In Wozniak (2013), a learning analytics approach was used to examine student engagement with online resources. The author found that providing presemester access and attending their first course in distance learning contributes to greater participation.

Lot of work about students' orientation and LA has been done by STELA (Successful Transition from Secondary to Higher Education through Learning Analytics) Erasmus+ project. In fact, according to De Laet, Broos, van Staalduinen, Leitner, and Ebner (2016), among the most common analytics items, academic achievement (e.g., credits obtained, GPA, timely graduation) is one of the most suitable indicators for a successful transition from secondary to higher education, and it can be characterized by different features. To enhance these measurables, Hodgson, Lam, and Chow (2010) suggest perceiving assessment during the transition as the basis of a critical environment in which students can develop confidence and become more sophisticated learners. This approach requires immediate and continuous feedback throughout the entire transition phase of each student. This feedback should not be summative, not connected to grades, but formative, so that students can monitor their learning in order to improve their learning process.

At the University of Turin, many of these strategies for preventing failure and improving success build their structure inside a DLE, even if most Italian universities do not consider digital transformation as one of the top three priorities. In fact, in OECD/European Union (2020), it emerges the present perspective about institutions' priorities in higher education in Italy. From the OECD Higher Education Institution Leader Survey to the question "Taking into account the HEInnovate dimensions/components listed below, please indicate the three that are most prominent in your strategy," none of the respondents considers the enhancement of digital transformation as one of the top three priorities for the institution. OECD/European Union (2020) also reports that the lowest number of investments in digital technologies and systems performed by Italian higher education institutions is devoted to developing digital skills of professors and researchers, who are the people directly in charge of students' education. It is worth mentioning that online universities did not participate to the survey.

In this context, in the subsequent paragraphs, we are going to describe the main orientation actions at the University of Turin adopting a DLE, addressing the increasing need of digitalization of higher education institutions.

One of the first actions carried out by the University of Turin is Orient@mente (https://orientamente.unito.it). Orient@mente consists of OOCs for the interactive exploration of courses and programs. Orient@mente helps in the preparation for admission tests, and in the recovery of gaps in many courses, from natural sciences to humanities. It serves as an effective dematerialized orientation for secondary school students who intend to enroll at university. This action provides three distinct types of intervention divided into specific areas:

- The "Realignment Courses" (RC) area, which delivers five OOCs in mathematics, biology, chemistry, physics, and Italian language to help students review, strengthen, and integrate the basic knowledge acquired during secondary school.
- The "Prepare for Tests" (PT) area, where students can assess themselves with numerous automatically evaluated tests that provide immediate and interactive feedback; this area helps students in preparing for admission tests to the various degree courses.
- The "Explore the Degree Programs of the University of Turin" (EDP) area, which consists of OOCs for the interactive exploration of the degree programs. Currently, over 80% of the degree programs at the University of Turin have an interactive orientation course on the platform.

Start@Unito (https://start.unito.it) promotes and facilitates the STT through the creation and dissemination of a series of OOCs, full university modules in many different first-year disciplines, covering topics that students will meet in most degree programs. Moreover, attending a full university module even before enrolling at the university helps students make a responsible choice about their academic career. Courses deliver interactive paths that facilitate university guidance and preparation for admission tests. Moreover, Start@Unito addresses other groups beyond students, namely, university professors and high school teachers. Considerably smaller than the traditional student population, both groups attended training sessions, and, apart from the action itself, the skills they acquired largely affect the secondary and higher education sector (Marchisio, Rabellino, Roman, Sacchet, & Salusso, 2019), especially in times of pandemic with forced online education (Bruschi, Marchisio, & Sacchet, 2021).

Foundation Programme (https://foundationprogramme.unito.it) facilitates the integration of pre-university students coming from countries where compulsory education lasts less than 12 years. These prospective students must attend an additional semester or a year and achieve the corresponding ECTS, according to the national standards. Because of its fully online setting, students can attend Foundation Programme from their home country, temporarily avoiding permissions, visa, or traveling. These students must only take the final exam in-person at the university, and they can do it even right before enrollment (Floris, Marchisio, Marello, & Operti, 2019).

5.3 Research Questions and Methodology

As evidenced by the literature (Barana et al., 2017; Barana, Bogino, Fioravera, Marchisio, & Rabellino, 2016a, 2016b, 2017), Orient@mente has a great impact on strategies to prevent student failure, both from the point of view of the learning experience, with automatic assessment and courses, and from the organizational point of view, with information on courses and programs and with interviews to older students or professors. The objective of this research is the evaluation of the impact of Orient@mente on students' path inside university.

The research question (RQ) is: How large is the impact of the orientation action of Orient@mente on students' grades at university?

Since the Orient@mente action takes place in a DLE, the environment collects data about the usage and the interaction of the users. The advantages of the Orient@mente action and the usage of the DLE are part of previous works; on the other hand, in this research we want to measure the correlation between academic grades and results of students who used Orient@mente, compared to students who did not use it. Since the orientation action aims at facilitating STT, every statistically significant improvement in admission tests and exams performed by students who used the platform is worth the effort of delivering such a large number of contents. The analysis considers data from two main sources: the database collecting careers of students enrolled at the University of Turin (restricted to the degree programs that have a course in the Orient@mente platform) and students' data from the interaction with the online platform. The latter kind of data are divided into two main sources: the first one collects students' interaction with the DLE, while the other data source is the database of the AAS, which contains everything about tests, such as grades, number of attempts, time, and other data.

The null hypothesis (H0) wonders if there is a significant difference in students' academic performance in relation to their activity on the Orient@mente platform.

We measured academic performance calculating the weighted average grade and the average number of ECTS acquired by university students. These values are the most relevant, since the university evaluation system uses similar indicators. Among other values, the ANVUR (Italian national agency for the evaluation of universities and research institutes) considers the number of students that gain more than 40 ECTS (indicator iA1 about didactics). This value is in fact an excellent aggregator for the dimensions of students who continue their studies and productivity. The analysis considers the previous education of students, such as the type of school attended and the graduation mark. Moreover, the analysis also takes into consideration the differences in users with respect to the area of Orient@mente that they attended.

5.3.1 Analysis Process

The whole process of collecting, managing, and analyzing data took place according to precise literature and indicators (Marchisio et al., 2019; Scheffel, Drachsler, Stoyanov, & Specht, 2014).

The data analysis was carried out following four different phases:

- Phase one (P1). The first action on the datasets was merging them to create a unique large database. The joining feature between the different databases lies in the email field, which is common through the different sources, but may anyway not be unique: users' data collected by the university contain two email addresses, the personal one and the institutional one.
- Phase two (P2). The second action on the merged dataset was cleaning and transforming. This phase gives an overview of the whole population and prepares the data for the further phases (grouping and statistical analysis).
- Phase three (P3). Statistical testing, with *t*-test, chi-square test, and ANOVA, to accept or reject the null hypothesis (H0).
- Phase four (P4). Group the items of the dataset according to variables expressing the use of the platform and which areas the student attended, in correlation with the student's achievement at university level.

Two software supported the four phases: "KNIME Analytics Platform 4.1.1" and "SAS 9.4." "KNIME," used in (P1) and (P4), allows easy access to different data sources, and, with the use of SQL language, it allows easy preparation of the dataset for next phases. In order to act on the merged dataset in phases (P2) and (P3), "SAS" allows cleaning, transforming the database and powerful statistical analysis. In this work, we will focus on the first three phases.

5.4 Results

The total sample has a size of 29,256 observations. Every record is a student who enrolled at the University of Turin from 2010 until the academic year 2017/2018. The dataset is divided into two subsamples: the first one (SS1) consists of students that are Orient@mente users too, 22.38% of the global sample, while the second one (SS2) consists of students who are not Orient@mente users, 77.62% of the sample. Table 5.1 describes the percentage of users divided into (SS1) and (SS2) and grouped by leaving certificate grade, for which the minimum grade is 60 and the maximum grade is 100:

- R1: First range, students with final examination grade between 60 and 69.
- R2: Second range, students with final examination grade between 70 and 79.
- R3: Third range, students with final examination grade between 80 and 89.
- R4: Fourth range, students with final examination grade between 90 and 100.

Table 5.1 Percentages and numbers of users according to the subsample and grouped by the final examination grade

Group	SS1	SS2	Row total
R1 students	1366 students	5023 students	6389 students 21.84%
Percentage	4.67%	17.17%	
Column percentage	20.86%	22.12%	
R2 students	2029 students	6851 students	8880 students 30.35%
Percentage	6.94%	23.42%	
Column percentage	30.99%	30.17%	
R3 students	1747 students	6081 students	7828 students 26.76%
Percentage	5.97%	20.79%	
Column percentage	26.68%	26.78%	
R4 students	1406 students	4753 students	6159 students 21.05%
Percentage	4.81%	16.25%	
Column percentage	21.47%	20.93%	
Column total	**6548 students** **22.38%** **100%**	**22,708 students** **77.62%** **100%**	**29,256 students** **100%**

Table 5.2 Percentages and numbers of users according to the subsample and grouped by gender

Gender	SS1	SS2	Row total
Male	3246 students	11,160 students	14,406 students
Row percentage	22.53%	77.47%	100%
Column percentage	49.58%	49.15%	49.24%
Female	3302 students	11,548 students	14,850 students
Row percentage	22.24%	77.76%	100%
Column percentage	50.42%	50.85%	50.76%
Column total	**6548 students**	**22,708 students**	**29,256 students**
Row percentage	**22.38%**	**77.62%**	**100%**

Table 5.1 shows that the distribution of students across the various ranges of grades is quite uniform, with percentages varying around 20 and 30%. Even inside every range, the global proportion between (SS1) and (SS2) is respected. A similar procedure, shown in Table 5.2, was carried out about gender, with no significant difference in the proportion of males and females.

As shown in Table 5.3, data proportion is different with respect to the educational origin. It appears that students from the "Lyceum" high school attended Orient@mente less. This happens because most of the OOCs available in Orient@mente until 2018 deal with scientific disciplines, which are the focus of this kind of school; thus students require less remedial intervention.

As for the students who used Orient@mente, first access inside the DLE of 68.10% of (SS1) students was prior to enrollment at the university. Most of the (SS1) students (79%) registered with the personal email address and did not use the university institutional email to access Orient@mente, even when the first access to

Table 5.3 Percentages and numbers of users according to the subsample and grouped by educational origin

Educational origin	SS1	SS2	Row total
International	85	124	209
Percentage	0.29%	0.42%	0.71%
Row percentage	40.67%	59.33%	
Column percentage	1.30%	0.55%	
Professional school	437	1314	1751
Percentage	1.49%	4.49%	5.99%
Row percentage	24.96	75.04%	
Column percentage	6.67%	5.79%	
Technical school	1638	4432	6070 20.75%
Percentage	5.60%	15.15%	
Row percentage	26.99%	73.01%	
Column percentage	25.02%	19.52%	
Lyceum	3977	14,922	18,899 64.60%
Percentage	13.59%	51.00%	
Row percentage	21.04%	78.96%	
Column percentage	60.74%	65.71%	
Unknown	411	1916	2327
Percentage	1.40%	6.55%	7.95%
Row percentage	17.66%	82.34%	
Column percentage	6.28%	8.44%	
Column total	**6548 students**	**22,708 students**	**29,256 students**
Row percentage	**22.38%**	**77.62%**	**100%**

the platform occurred after enrollment (16%). Thus, in phase (P1) some (SS1) students did not appear in the merged dataset: they might have used two different personal emails, one for Orient@mente and one to enroll at the University of Turin. Such cases cannot be traced back to the university student and this item will appear in (SS2) in the dataset.

5.4.1 Hypothesis Tests

The hypothesis test checks the equality between the average number of ECTS (which we will indicate with the Greek letter μ) achieved by the students at the end of the first academic year in the subsamples (SS1) and (SS2). The test is one-tailed; the alternative hypothesis entails just a larger average for (SS1) students:

$$H_0 : \mu_{\mathrm{SS1}} = \mu_{\mathrm{SS2}}$$

$$H_A : \mu_{\mathrm{SS1}} > \mu_{\mathrm{SS2}}$$

The *p*-value for the test is much less than 0.0001, thus rejecting the hypothesis of equality with high statistical significance. We can state that there is statistical difference between the average number of ECTS achieved by students depending on whether they logged in the Orient@mente DLE.

The same hypothesis test checked the equality of the weighted average grade of students between the subsamples (SS1) and (SS2). The *p*-value for the test is almost null; thus we can reject the hypothesis of equality with high statistical significance. Even in this second case, Orient@mente action proved effective.

The various groups of students can bring different features on the global behavior according to the same measurable. We restricted the analysis on the sample considering the four ranges of leaving certificate grade (R1), (R2), (R3), and (R4). A significant difference between the average number of ECTS appears in the comparison of (SS1-R1), (SS2-R1) and of (SS1-R4), (SS2-R4) with *p*-value less than 0.005. On the other hand, for groups (R2) and (R3), we cannot reject the null hypothesis; equality is possible. In particular, the case of (R3) is quite significant, since difference between the two averages (SS1-R3) and (SS2-R3) is very little, below 0.05. Thus, we can infer that Orient@mente action is more effective for users with the high-performing and low-performing students, according to the final examination grade. These two groups contain two completely diverse kinds of students. With a closer look at the behavior of students in (R1) and (R4), we can notice that 35% of (SS1-R1) students logged in after enrollment, while only 25% of (SS1-R4) students did it. This shows that weaker students use Orient@mente DLE for one of its main purposes, to fill the gaps that emerge, even during the first year of university studies.

Given the significant difference between the average of the ECTS between the two samples SS1 and SS2, we performed a more in-depth analysis on courses' subscriptions of Orient@mente users to understand which combination of activities within the platform is more significant.

We defined a new variable called "areas used," assigned to SS1 users to indicate which of the areas of the Orient@mente platform each student subscribed to. We marked records with one or more of the following labels:

- PT for a subscription to at least one course in the area "Prepare for Tests".
- RC for a subscription to at least one course in the area "Realignment Courses".
- EDP for a subscription to at least one course in the area "Explore the Degree Programs of the University of Turin".
- Nothing for SS2 students.

Table 5.4 shows the distribution of the variable ECTS among the various categories.

Table 5.4 provides a first look at the most helpful areas. Looking at the single areas, users of EDP gained the largest number of ECTS on average. This is in line with the purposes of Orient@mente: with proper guidance in the choice of future studies, the students can perform better during exams. The reader should keep in mind that it is possible that EDP users can be students who have clear ideas for their future and just subscribed to one course to confirm their existing opinion. On the

Table 5.4 Descriptive analysis of the distribution of the average of ECTS achieved by the students at the end of the first academic year with respect to the values of the "areas used" variable

Areas used	Min	Mean	Max	IQR
Nothing (SS2 students)	0	28.88	116	41
PT	0	30.75	119	41
RC	0	27.12	85	37
EDP	0	31.49	116	42
RC + EDP	0	28.53	73	41
RC + PT	0	26.31	97	39
RC + PT + EDP	0	30.54	82	46
PT + EDP	0	32.72	102	40

other side, the area with less ECTS on average is RC, and this is not surprising because this area targets students with more difficulties, also students with additional formative duties.

In order to evaluate the differences, after verifying the hypotheses of applicability, we performed an ANOVA test to compare the average ECTS between these groups. The p-value of the test is much lower than 0.0001 and this leads us to reject the null hypothesis; there is no equality between the means of the various groups. Analyzing more in depth the internal difference between the two groups, we performed a multiple comparison test with the Bonferroni correction to analyze in which cases the difference in the average number of ECTS is significant. The test showed a statistically positive difference between the following couples:

- PT and "Nothing".
- PT + EDP and "Nothing".
- PT and RC + PT.
- PT + EDP and RC.

This result shows how the "Prepare for Tests" area has a large weight in the preparation of students since it always appears as positively higher than other or no areas. From these significant tests, we can infer that PT and EDP are the areas with larger impact, more than RC, which is dedicated to students with more troubles in recovering gaps in the knowledge.

5.4.2 Focusing on Two Courses

We have already seen a dependence of the effectiveness of Orient@mente on the final examination grade. Other differences arise when dealing with students enrolled in different degree programs. We are going to focus on two cases with conflicting behavior.

Case 1 (C1) concerns students in mathematics (1271 students). In (C1), we detect an opposite behavior with respect to the global sample: the average number of

ECTS earned by (SS2) students is even 3.5 points higher (p-value <0.02) than (SS1) students. This is not surprising: a possible explanation could be a greater awareness of students in Mathematics, as this is a well-known topic, and the university choice could be guided by an established knowledge of the intrinsic difficulties of a STEM learning path. Usually only the best school students in Mathematics choose this topic, so they are typically students with less gaps to fix.

On the other hand, case 2 (C2) concerns students in biological science (2041 students). In (C2) the global behavior is emphasized: the average number of ECTS achieved by (SS1) students is 4 points higher with respect to (SS2) students (p-value <0.0001). In many degree programs, four ECTS mean one extra exam during first year. In such programs, modules like mathematics, physics, and chemistry are not the main topics but basic disciplines and often represent first-year obstacles. They are the first modules that students encounter, but not the main discipline of the program.

We understood that focusing on the conflicts can lead to interesting phenomena, but these are not supported by specific evidence, and the analysis should leave the quantitative method in favor of a qualitative one, by using social analysis to discover the real motivation behind these numbers.

We should investigate further to track down which user data can be acquired to enable a clear evaluation of these motivations, so as to better support student in the best choice for their university life.

5.5 Conclusions

According to the analysis on students' data, it emerges that the effectiveness of Orient@mente has statistical significance, thus proving that it positively affects students' orientation and first-year outcomes, thanks to the OOC, accessible even before enrolling at the university (RQ1). In this research, we described a method that can analyze in the same way the effectiveness of similar experiences. The method could be refined in order to automatically execute and return real-time results.

Based on our experience, we can list some recommendations for a digital intervention on university orientation:

- Deliver a system in agreement with the other orientation offices and activities, to maximize the positive effect.
- Let students monitor and track their current progress and knowledge with automatic tests and, more importantly, feedback.
- Develop the system in such a way that it can answer to different needs, adapting the learning activities according to the students' pace and needs.
- Provide students an environment that can help them recover their gaps.

These results guide future orientation actions. Given the results, the analysis, and the evaluation of the impact of Orient@mente, the University of Turin immediately took action and decided to invest in this initiative. New Orient@mente OOCs and new test areas cover more than 85% of the university degree programs.

Moreover, the model of Orient@mente is going to be exported in other universities (Marchisio, Margaria, & Sacchet, 2020).

Many more parameters guide the evaluation process over the guidance service. Close future research concerns:

- Students' personal profile fields, to provide a personalized orientation experience, suited to individual requirements, to offer students the content they need most when they need it, according to their online activity.
- Design personalized learning paths using machine learning techniques to differentiate students' learning paths.
- Cluster analysis to identify "behavior" patterns within the platform and study their effects. In this way, once recognized nonproductive behaviors by new users, it is possible to intervene to suggest a more effective learning path.
- Learning analytics techniques to connect the results of tests carried out with the automatic assessment system to improve the implementation of automatic formative assessment strategies and provide immediate and personalized feedback to students (Barana et al., 2019).
- Analysis of other variables, such as age, region of origin, results of the admission test, and chosen program.
- The comparison of Orient@mente OOCs quality with the international benchmark and with the standards of MOOC providers (Marchisio & Sacchet, 2020).

Acknowledgments The authors would like to thank the bank foundation Compagnia di San Paolo, which financially supports many of the initiatives at the University of Turin, in particular the project OPERA, Open Program for Educational Resources and Activities, in which this research took place. The authors would like to thank Dr. Massimo Bruno, Director of the Didactics Office at the University of Turin, and his staff for the support they gave to the action Orient@mente.

References

Barana, A., Bogino, A., Fioravera, M., Marchisio, M., & Rabellino, S. (2016a). Digital support for university guidance and improvement of study results. *Procedia—Social and Behavioral Science, 228*, 547–552. https://doi.org/10.1016/j.sbspro.2016.07.084

Barana, A., Bogino, A., Fioravera, M., Marchisio, M., & Rabellino, S. (2016b). Open platform of self-paced MOOCS for the continual improvement of academic guidance and knowledge strengthening in tertiary education. In *Extended Abstracts of the Multiconference* (pp. 14–25) Genova University Press. https://doi.org/10.20368/1971-8829/1383

Barana, A., Bogino, A., Fioravera, M., Floris, F., Marchisio, M., Operti, L. & Rabellino, S. (2017a). Self-paced approach in synergistic model for supporting and testing students: The transition from Secondary School to University. In *Proceedings of 2017 IEEE 41st Annual Computer Software and Applications Conference (COMPSAC)* (pp. 404–409). IEEE. https://doi.org/10.1109/COMPSAC.2017.211

Barana, A., Bogino, A., Fioravera, M., Marchisio, M., & Rabellino, S. (2017b). Open platform of self-paced MOOCs for the continual improvement of academic guidance and knowledge strengthening in tertiary education. *Journal of E-Learning and Knowledge Society, 13*(3). https://doi.org/10.20368/1971-8829/1383

Barana, A., Conte, A., Fissore, C., Marchisio, M., & Rabellino, S. (2019). Learning analytics to improve formative assessment strategies. *Journal of E-Learning and Knowledge Society, 15*(3), 75–88. https://doi.org/10.20368/1971-8829/1135057

Bruschi, B., Marchisio, M., & Sacchet, M. (2021). Online teaching in higher education with the support of Start@Unito during Covid-19 pandemic. In L. S. Agrati et al. (Eds.), *Bridges and mediation in higher distance education* (HELMeTO 2020. Communications in computer and information science) (Vol. 1344). Springer. https://doi.org/10.1007/978-3-030-67435-9_15

De Laet, T., Broos, van Staalduinen, J. P., Leitner, P., & Ebner, M. (2016). Successful transition from secondary to higher education using learning analytics. In *Workshop of the 44th SEFI Conference, 12–15 September 2016, Tampere, Finland*.

DEAP. (2020). *Digital education action plan 2021–2027—Communication*. Retrieved July 29, 2021, from https://ec.europa.eu/education/sites/default/files/document-library-docs/deap-communication-sept2020_en.pdf

Early leavers from education and training. (2021). Retrieved July 3, 2021, from https://ec.europa.eu/eurostat/statistics-explained/index.php/Early_leavers_from_education_and_training

Educational attainment statistics. (2021). Retrieved July 3, 2021, from https://ec.europa.eu/eurostat/statistics-explained/index.php/Educational_attainment_statistics

European Commission. (2010). *EUROPE 2020 A strategy for smart, sustainable and inclusive growth*. Retrieved July 3, 2021, from https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52010DC2020&from=en

Floris, F., Marchisio, M., Marello, C. & Operti, L. (2019). Bridge the gap between high school systems with less than twelve years of schooling and European Universities. In *Proceedings of the 5th International Conference on Higher Education Advances (HEAd'19), Valencia, Spain* (pp. 1207–1216). https://doi.org/10.4995/HEAd19.2019.9494.

Floris, F., Marchisio, M., Rabellino, S., & Sacchet, M. (2020). Learning analytics to evaluate the effectiveness of higher education student failure prevention. In *Proceedings of the 17th International Conference on Cognition and Exploratory Learning in the Digital Age (CELDA 2020)* (pp. 265–272).

Gašević, D., Dawson, S., & Siemens, G. (2015). Let's not forget: Learning analytics are about learning. *Tech Trends, 59*, 64–71. https://doi.org/10.1007/s11528-014-0822-x

Hajian, S., Bonchi, F., & Castillo, C. (2016). Algorithmic bias: From discrimination discovery to fairness-aware data mining. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16)* (pp. 2125–2126). Association for Computing Machinery. https://doi.org/10.1145/2939672.2945386.

Hodgson, P., Lam, P., & Chow, C. (2010). In *International Conference on Assessment Experience of First Year University Students: Dealing with the Unfamiliar, Enhancing Learning Experiences in Higher Education*

Krull, G., & Duart, J. (2018). Meeting the needs of digital learners: Learner support patterns and strategies. In: *13th International Conference on E-Learning (ICEL 2018), Cape Town, South Africa*.

Lonn, S., Aguilar, S. J., & Teasley, S. D. (2015). Investigating student motivation in the context of a learning analytics intervention during a summer bridge program. *Computers in Human Behavior, 47*, 90–97. https://doi.org/10.1016/j.chb.2014.07.013

Marchisio, M., Margaria, T., & Sacchet, M. (2020). Automatic formative assessment in computer science: Guidance to model-driven design. In *Proceedings of 2020 IEEE 44th Annual Computer Software and Applications Conference (COMPSAC)* (pp. 201–206). https://doi.org/10.1109/COMPSAC48688.2020.00035.

Marchisio, M., Rabellino, S., Roman, F., Sacchet, M., & Salusso, D. (2019). Boosting up data collection and analysis to learning analytics in open online contexts: An assessment methodology. *Journal of E-Learning and Knowledge Society, 15*(3), 49–59. https://doi.org/10.20368/1971-8829/1135048

Marchisio, M., & Sacchet, M. (2020). Analysis items to assess the quality of open online courses for higher education. In *Proceedings of International Conference E-Learning 2020* (pp. 63–70). https://doi.org/10.33965/el2020_202007L008.

Marquez-Vera, C., Morales, C. R., & Soto, S. V. (2013). Predicting school failure and dropout by using data mining techniques. *IEEE Revista Iberoamericana de Tecnologias del Aprendizaje, 8*(1), 7–14. https://doi.org/10.1109/RITA.2013.2244695

Ochoa, X., & Merceron, A. (2018). Quantitative and qualitative analysis of the learning analytics and knowledge conference 2018. *Journal of Learning Analytics, 5*(3), 154–166. https://doi.org/10.18608/jla.2018.53.10

OECD/European Union. (2020). *Supporting entrepreneurship and innovation in higher education in Italy. OECD skills studies*. OECD Publishing. https://doi.org/10.1787/43e88f48-en.

Riazy, S., Simbeck, K., & Schreck, V. (2020). Fairness in learning analytics: Student at-risk prediction in virtual learning environments. In *Proceedings of the 12th International Conference on Computer Supported Education* (Vol. 1, pp. 15–25). CSEDU. https://doi.org/10.5220/0009324100150025.

Scheffel, M., Drachsler, H., Stoyanov, S., & Specht, M. (2014). Quality indicators for learning analytics. *Educational Technology & Society, 17*(4), 117–132. Retrieved from http://www.jstor.org/stable/jeductechsoci.17.4.117

Schumacher, C., & Ifenthaler, D. (2018). The importance of students' motivational dispositions for designing learning analytics. *Journal of Computing in Higher Education, 30*, 599–619. https://doi.org/10.1007/s12528-018-9188-y

Wozniak, H. (2013). Student engagement with an online orientation resource: How learning analytics refines educational design principles. In J. Herrington, A. Couros & V. Irvine (Eds.), *Proceedings of EdMedia 2013—World Conference on Educational Media and Technology* (pp. 941–950). Association for the Advancement of Computing in Education (AACE). Retrieved May 6, 2021, from https://www.learntechlib.org/primary/p/112074/ (last accessed July 3, 2021).

Chapter 6
Examining the Influence of Cognitive Ability on Situating to a Video Game: Expanded Discussion

Michael P. McCreery, Danielle L. Head, Sam A. Leif, Joseph P. Fiorentini, Catherine A. Bacos, Jeffrey R. Laferriere, S. Kathleen Krach, and Le Quanda L. Cole

6.1 Introduction

Individuals vary in their abilities to process information, construct meaning from that information, and apply it to real-world situations (Glaser, 1985; Holzman, Pellegrino, & Glaser, 1983). Individual variations in these skills are broadly identified as differences in a person's cognitive ability. Cognitive differences within an individual have demonstrated outcome differences in a variety of contextualized experiences (Goh, Truman, & Barber, 2019; Jonassen & Grabowski, 2012). Video games can provide one such contextualized experience (Gee & Gee, 2017; Squire, 2006), combining content and narrative to create a complex system in which individual differences in cognitive skill have the potential to impact learning (Schrader, McCreery, Carroll, Head, & Laferriere, 2019).

Other intra-individual variations have the potential to differentially shape contextualized experience outcomes. For example, differences in video game expertise impact in-game experiences (McCreery, Schrader, & Krach, 2011). When compared to experts, video game novices demonstrate difficulty with spatial navigation (Murias, Kwok, Castillejo, Liu, & Iaria, 2016; Ventura, Shute, Wright, & Zhao, 2013), navigating the graphical user interface (GUI; Cockburn, Gutwin, Scarr, &

The original version of this chapter was revised. The correction to this chapter is available at https://doi.org/10.1007/978-3-030-90944-4_17

M. P. McCreery (✉) · D. L. Head · S. A. Leif · J. P. Fiorentini · C. A. Bacos · L. Q. L. Cole
College of Education, University of Nevada, Las Vegas, Las Vegas, NV, USA
e-mail: michael.mccreery@unlv.edu

J. R. Laferriere
Lebanon Valley College, Annville, PA, USA

S. Kathleen Krach
Florida State University, Tallahassee, FL, USA

D. Ifenthaler et al. (eds.), *Orchestration of Learning Environments in the Digital World*, Cognition and Exploratory Learning in the Digital Age,
https://doi.org/10.1007/978-3-030-90944-4_6

Malacria, 2015), and prioritizing information (Yan, Huang, & Cheung, 2015). The situated understanding of the game and its internal systems play a significant role in how contextualized experience manifests (McCreery et al., 2011). Therefore, in order to use video games as research contexts, players must be able to discern how to use the system on a fundamental level. In particular, intra-individual differences are prevalent in the acclimation stage, during which people become oriented to and within a domain (Alexander, Sperl, Buehl, Fives, & Chiu, 2004).

McCreery et al. (2011) state that an acclimation stage must be considered in any future research when both experts and novices are included in the sample. In addition, they called for further research into how other intra-individual differences (e.g., cognitive ability) may shape the nature of the acclimation stage (McCreery et al., 2011). The current study was designed in response to this call with the goal of exploring the role of cognitive ability in shaping how players think within, and acclimate to, video games as complex, contextualized systems.

6.2 Theoretical Framework

Complex systems are defined as a multiplicity of components that interact with one another in dynamic and emergent ways, resulting in the production of complex behaviors (Hilpert & Marchand, 2018). By design, most video games can be described as complex systems (Tekinbaş & Zimmerman, 2004). Video games consist of a multiplicity of components that include, but are not limited to, dimensional space, a graphical user interface (GUI), and in-game objects. Each of these components interacts with one another and with the player in a dynamic manner. Finally, the player responds to these dynamic interactions by producing complex in-game behaviors (e.g., interacting with other characters, making choices, etc.) As these interactions occur over time, emergent properties associated with game design shape subsequent player behavior (Tekinbaş & Zimmerman, 2004). For these reasons, a player must possess a basic level of skill and understanding within this complex system in order to attend to and accurately interpret the feedback provided by interactions with the rules, goals, and objectives (McGonigal, 2011). These complex system characteristics interact with individual user differences in a cyclical manner, whereby game structures shape player behavior and player behavior shapes game structures (McCreery, Vallett, & Clark, 2015).

6.2.1 Individual Differences: Expertise

Novice and expert video game players exhibit different behaviors in-game that can be attributed to their accumulated experience (Murias et al., 2016; Ventura et al., 2013; Yan et al., 2015). Ventura et al. (2013) found that video game use was significantly and positively correlated with in-game spatial navigation. This finding was extended to show that prior video game use was also associated with better in-game controls proficiency and more efficient orientation and navigation strategies (Murias et al., 2016). Expert play has also been linked to increased efficacy of task switching, interface control, and time-pressured strategic decision-making in-game when compared to novice play (Yan et al., 2015). Cockburn et al. (2015) argued that performance with GUIs is contingent on prior experience as well. Novice users tend to utilize comparatively inefficient visual search strategies during GUI navigation and lack strategic knowledge that facilitates mapping the system tools to in-game tasks (Cockburn et al., 2015).

Considering such complexity, researchers have sought to understand the implications that individual differences have on research and instruction within video games. For example, while examining differences among experts and novices within the game *World of Warcraft*, time-series data revealed that a 30-min acclimation stage was warranted for players who had never before played the game (McCreery et al., 2011). Specifically, novices demonstrated difficulty interacting with in-game objects, as well as recognizing environmental and graphical user interface cues that were critical to learning. For these reasons, McCreery et al. (2011) stress the importance of an orientation period. This orientation period, grounded in the principle of acclimation (Alexander et al., 2004), is of chief importance to new players; it allows players to orient themselves to the game space, navigate the GUI, and prioritize information about presented feedback. Without this time to acclimate, players may miss important content intended to further game goals and objectives, which can negatively impact their ability to succeed in the game (McCreery et al., 2011).

6.2.2 Individual Differences: Cognitive Ability

Exploring intra-individual differences can provide context for conceptualizing the manner in which players interact with video games. One such intra-individual difference that may influence how the acclimation stage emerges is cognitive ability. Stemming from the work of Cattell (1971, 1987), cognitive ability consists of two components: *crystallized intelligence* and *fluid intelligence*. As Zaval and colleagues (2015) explain, crystallized intelligence is characterized as knowledge that has been acquired across the lifespan; it reflects an amalgamation of experience, culture, and prior learning, gained through negotiating a variety of tasks, situations, and challenges. They further explain that fluid intelligence is characterized by a broad range of skills including processing speed, attention, and working memory. Zaval and team (2015) posit that these skills are intended to produce, change, and manipulate novel information in real time. Crystallized intelligence is typically measured using more generalized knowledge-based assessments, while fluid intelligence is typically measured using novel stimuli tasks.

Given video game complexity and cyclical nature, researchers have long been intrigued by the relationship between cognitive ability and video games. Currently, research has clustered around three perspectives: (1) video games as valid psychometric measures of cognitive ability (Buford & O'Leary, 2015; Quiroga, Diaz, Román, Privado, & Colom, 2019), (2) video games as cognitive training tools to influence cognitive ability (Bediou et al., 2018; Fikkers, Piotrowski, & Valkenburg, 2019), and (3) cognitive ability as an influence on video game performance (Kokkinakis, Cowling, Drachen, & Wade, 2017; Rabbitt, Banerji, & Szymanski, 1989). In the following sections, the chapter will examine each of these areas of study in further depth.

6.2.3 Video Games as Psychometric Measures of Cognitive Ability

The relationship between video games and cognitive factors has been examined for over three decades, with robust correlations aligning video game performance and tests of intelligence (Jones, Dunlap, & Bilodeau, 1986; Rabbitt et al., 1989). As early as 1986, Jones and colleagues made use of off-the-shelf Atari video games available at the time to investigate whether game-based tasks (e.g., reaction time) shared variance with existing cognitive tests. They found that video games had much in common with cognitive tests, but also contained substantial amounts of variance not shared with cognitive factors (Jones et al., 1986). This work was followed by the development of a video game specifically for research purposes, *Space Fortress* (Mané & Donchin, 1989). This game exhibited correlations as high as $r = 0.68$ between a validated intelligence test and participants' in-game scores after training (Rabbitt et al., 1989). The strength of that relationship was comparable to concurrent validity investigations of paper-and-pencil cognitive tests available at the time, $r = [0.58–0.85]$ (Heim, 1968). The promise of these initial findings led to further development of video games as direct adaptations of cognitive subtests.

For example, the Digit Symbol subtest of the Wechsler Adult Intelligence Scale, Third Edition (WAIS-III; Wechsler, 1997) was adapted as a spaceship shooter entitled *Space Code*. When compared to its paper-and-pencil equivalent, the game exhibited a correlation of $r = 0.64$ (McPherson & Burns, 2007). This was followed by the development of *Space Matrix* (McPherson & Burns, 2008), an extension of *Space Code* that added a simultaneous task to the game that was based on the Dot Matrix task (Miyake, Friedman, Rettinger, Shah, & Hegarty, 2001). This adaptation also correlated strongly, at $r = 0.66$, with its originating psychometric test. These findings support the development and use of video games that align to task-based measures of intelligence. However, the time and expense of designing game-based measures of cognitive tasks proved to be prohibitive, leading to increased interest in the ability of off-the-shelf games to measure intelligence constructs (Quiroga et al., 2019).

Recognizing the potential of off-the shelf games, Baniqued et al. (2013) adopted a broad-reaching mixed-methods approach to qualitatively classify and quantitatively investigate the relationship between a number of casual video games (e.g., *Bejeweled*, *Solitaire*, *Minesweeper*) and a swath of cognitive skills. The principal components of working memory and spatial reasoning created from game-based scores were found to correlate at $r = [0.55–0.65]$ with principal components of working memory and fluid intelligence created from the intelligence task scores (Baniqued et al., 2013). These findings narrowed the focus to commercially available "brain games" that were specifically selected to represent intelligence-related abilities (e.g., memorization, visualization, analyzation, computation) (Quiroga et al., 2015). This approach led to a correlation between the latent factors of video game performance and general intelligence of $r = 0.93$ and the ability to near-perfectly predict the general intelligence score from performance on the selected games, $r^2_{\text{adusted}} = 0.96$ (Quiroga et al., 2015). Quiroga et al. (2015) note, however, that modification of commercial games to allow control (e.g., score criteria, difficulty) and consistency (e.g., order of presentation, display) in lab-based research is recommended.

To that end, the relationship between video games and fluid intelligence was further strengthened by analysis investigating *Portal 2*, a puzzle-platforming game, as a valid and reliable measure of fluid intelligence (Foroughi, Serraino, Parasuraman, & Boehm-Davis, 2016). A test battery was constructed within the game using Puzzle Creator that built progressively more difficult chambers for participants to navigate as a measure of fluid intelligence. Scores from the game showed an internal consistency reliability of $\alpha = 0.80$ and a concurrent validity correlation of $r = 0.65$ when compared to a paper-based test of inductive reasoning (Foroughi et al., 2016). Video game performance and cognitive factors appear to be intrinsically tied, a premise which has guided the use of video games as psychometric measures of intelligence. It has also guided the idea that video games may be used as training tools or cognitive-based interventions to influence intelligence scores.

6.2.4 Video Games as Cognitive Training Instruments

Interest in enhancing cognitive ability through brain training has been investigated for decades. Early findings examining working memory supported the claim that cognition is highly malleable when exposed to training (Chase & Ericsson, 1982). This led to a series of studies that supported the effect of cognitive gains across two tightly linked domains but did not support the effect of gains across loosely related domains (Sala, Tatlidil, & Gobet, 2018). Cognitive training, whether task specific or broader in scope, only seems to result in improvement in trained skills and results in no significant effects on non-trained skills such as problem-solving and general reasoning (Sala et al., 2018).

However, in contrast to these findings, video game training has generated extensive interest as a tool for cognitive enhancement, supported by several studies that showed early evidence of a causal relationship between action video game training and cognitive ability improvement (Bejjanki et al., 2014; Green & Bavelier, 2003). Bolstered by the framework of "learning to learn," it is theorized that action video game players gain experience in making decisions based on limited, noisy information that occurs during video gameplay (Bavelier, Green, Pouget, & Schrater, 2012). This experience, in turn, has shown increased performance on a broad range of cognitive tasks (e.g., visual short-term memory, reaction time, attention, etc.), a finding which runs counter to the cognitive training literature (Bavelier et al., 2012). The tasks used to compare the cognitive ability of video game players and non-video game players in these studies (e.g., locating a target among distractors) can be framed as requiring probabilistic inference or determining the most likely choice given a set of limited task information (Bavelier et al., 2012). Action video game players evidence better abilities to process task-relevant information and exclude task-irrelevant information. This skill can be trained in non-video game players through action video gameplay and may act as a transferable general learning mechanism (Bejjanki et al., 2014).

This pattern of results was also found when examining visual attention of action video game players; video game players outperformed non-video game players on multiple aspects of visual attention (e.g., attentional capacity, visuomotor coordination), and action video game training resulted in improved performance (Green & Bavelier, 2003). Nonaction video gameplay has also been linked to cognitive benefits. For example, *Tetris*, a puzzle game, was used as a training tool to improve mental rotation and spatial visualization performance (Okagaki & Frensch, 1994), *Super Tetris* was used to improve reaction times (Goldstein et al., 1997), and *Rise of Nations*, a real-time strategy game, was associated with improvements in mental rotation and task switching (Basak, Boot, Voss, & Kramer, 2008).

Although these findings yielded a set of positive results, the body of literature suffers from replicability issues (Sala et al., 2018). Action video game training was not found to substantially improve performance for non-video game players on a series of cognitive transfer tasks in a longitudinal study that varied both practice schedule and training game (Boot, Kramer, Simons, Fabiani, & Gratton, 2008). Additionally, studies using nonaction video games have given rise to similar null findings. The effect of *Tetris* as a video game training tool for mental rotation improvement was found to disappear over repeated testing, as the control group made similar performance improvements reportedly due to repeated test exposure (Terlecki, Newcombe, & Little, 2008). The use of a real-time strategy game, *Starcraft: Brood War*, as an active training regimen did not yield support for transfer related to cognitive measures of processing speed, working memory, or reasoning (Minear et al., 2016).

Even the theoretical underpinnings of "learning to learn" have been challenged, with results suggesting the near-transfer effect of specific cognitive processes frequently used during video gameplay only improves cognitive performance in tasks that have common underlying demands, not general cognitive ability (Oei & Patterson, 2013). The relationship between video gameplay and its impacts on cognitive ability is still an active area of development for the field, but the reverse is of particular interest for the current chapter; what is the nature of cognitive ability's impact on in-game behaviors and outcomes?

6.2.5 Cognitive Ability as a Predictor of Video Game Outcomes

The influence of cognitive factors on in-game behaviors has also been an active, if sporadic, area of interest. In 1989, Rabbitt and colleagues found that intelligence scores on a paper-based general intelligence test predicted an individual's performance on video game tasks. After this initial investigation, there was an extended gap in practiced performance and task-based literature. Due in large part to the popularity and dominant interest of using games either as training tools or as psychometric measures, examinations of predictive relationships associated with intelligence and gameplay seemingly stagnated. However, there has been a recent revitalization of interest in the topic at the factor (e.g., fluid intelligence) level.

At the time of this writing, no literature was found that explicitly and directly used the construct of crystallized intelligence to examine its predictive relationship to video game performance. However, literature exists that indirectly aligns with the above definition of crystallized intelligence (i.e., knowledge acquired through education and experience). For example, research examining previous video game experience and performance demonstrated significant differences in how experienced players navigate the GUI and associated information within the game *Defense of the Ancients 2* (*DOTA 2*; Castaneda, Sidhu, Azose, & Swanson, 2016). For example, the more experienced the player, the less effort was spent monitoring in-game tools, and the more automated usage became. Alternatively, existing research identified fluid intelligence as a direct predictor of performance with video games (Baniqued et al., 2013; Kranz, Baniqued, Voss, Lee, & Kramer, 2017). For example, fluid ability was found to be positively correlated with video game performance across a variety of multiplayer online battlefield games (e.g., *League of Legends*, *DOTA 2*; Kokkinakis et al., 2017). In discussions associated with this correlation, the researchers raise the possibility "that rather than games modifying cognition, learning to play video games depends on the same cognitive resources underlying performance on intelligence tests" (Kokkinakis et al., 2017, pp. 7–8).

Thus, the field's current understanding of the interplay among cognitive resources and dynamic, emergent gameplay suggests that reframing how we examine video game performance is warranted. As such, the current investigation utilized these two overarching questions as guides: (a) how do individual differences in cognitive ability impact outcomes associated with a video game; and (b) if players were given the opportunity to play the game once as an acclimation activity and a second time as a skill demonstration, would there be changes in the predictive nature of cognitive ability on game outcomes? With these ideas in mind, the following research questions were designed:

RQ1. Does cognitive ability (Shipley-2 scores) predict gameplay outcomes for the first attempt of a new game (The Deed)?

RQ2. Does cognitive ability (Shipley-2 scores) predict gameplay outcomes for the second attempt of the same game?

RQ3. Does cognitive ability (Shipley-2 scores) predict the learning growth (outcome change) scores between gameplay one and two?

6.3 Method

6.3.1 Subjects

One hundred and forty-four (*N* = 144) participants completed the study (see Table 6.1). The gender breakdown included 39 males and 105 females. The racial makeup of the sample was as follows: 47% White, 7% Black or African American, 17% Hispanic or Latino, 16% Asian, 2% Native Hawaiian or Pacific Islander, 10% two or more races, and 1% other. The mean age was approximately 24 (SD = ~6) years.

6.3.2 Procedures

6.3.2.1 Game Selection

The Deed (Grab the Games, 2015) was selected as part of a more extensive study examining forced-choice games as a model for stealth assessment. *The Deed* is an appropriate choice for studying the effect of cognitive ability on gameplay outcomes. Firstly, the entire game can be completed in 20–30 min, falling within the acclimation time period (approximately 30 min) outlined by McCreery et al. (2011). Secondly, success in the game is predicated on the player's ability to interact with both design components and narrative. The game is considered a reverse murder mystery with problem-solving elements, in which the desired outcome is to gather information (i.e., evidence and weapons) to get away with a crime, rather than gathering information to solve a crime. If players are successful, they evade prison by implicating another character for their crime and may even receive the family inheritance. If they are unsuccessful, they are sentenced to prison once their crime is discovered.

Table 6.1 Demographics

	n	%
Race		
White	68	47.2
Black or African-American	10	6.9
Hispanic or Latino	24	16.7
Asian	23	16.0
Native Hawaiian or Pacific Islander	3	2.1
Two or more races	15	10.4
Other	1	0.07
Gender		
Male	39	27.1
Female	105	72.9

6.3.2.2 Data Collection

Data were collected during the 2017–2018 academic year from subject pool participants within a college of education in the southwestern United States. Participants were asked to complete a demographic questionnaire, followed by a first playthrough of *The Deed* (Grab the Games, 2015) video game. Participants were not provided verbal gameplay directions; rather the game provided a built-in introduction prior to gameplay. Following the first playthrough, participants completed the Shipley-2 and then completed a second playthrough of *The Deed*. All gameplay was captured for later coding and analysis using Fraps, a DirectX Windows-based application that allows for audio and video recording directly from the video card at up to 7680 × 4800 resolution and 120 frames per second (Version 3.5.99; Beepa Pty Ltd., 2013). The summative outcome was determined by reviewing the game capture footage.

6.3.3 Instruments

6.3.3.1 The Deed Outcome Data

Outcome data were calculated by creating a dichotomous score based on video observations of the players' activity. Specifically, overall outcome scores were coded as failure (prison and no inheritance = 0) or success (no prison, with or without inheritance = 1). Further, learning growth scores were coded as nongrowth (do the same or worse from Playthrough One to Playthrough Two = 0) or growth (do better on Playthrough Two compared to Playthrough One = 1).

6.3.3.2 Shipley-2

The Shipley-2 (Shipley, Gruber, Martin, & Klein, 2009) is a brief measure of crystallized (e.g., prior knowledge) and fluid intelligence (e.g., problem-solving). Subtest scores are provided for Vocabulary (crystallized intelligence), Abstraction (fluid intelligence), and Block Patterns (fluid intelligence). Composite scores are labeled as Composite A (Vocabulary with Abstraction) and Composite B (Vocabulary with Block Patterns). Composites A and B also each provide an Impairment Index score (not used in this study). Reviews for the Shipley indicate solid reliability and validity data to support its use (Hayes & Thorndike, 2010).

6.4 Results

Prior to analyses, the Box-Tidwell (1962) test was conducted to check that the assumption of linearity was upheld. For all analyses, interaction terms were found to be nonsignificant, $p_{min} = 0.092$. The Shipley-2 subscales were used as continuous predictor variables (see Table 6.2). Gameplay outcome was coded as a dichotomous nominal variable, with "0" denoting a failed outcome and "1" denoting a successful outcome. Gameplay growth was also coded as a dichotomous nominal variable, with "0" denoting nongrowth and "1" denoting growth between the first and second playthrough outcomes. Although logistic regression allows for the classification of outcomes, that was not the purpose of this study.

6.4.1 RQ1

Analyses containing cognitive ability should maintain the structure of the composite scales (Sattler, 2008). Therefore, two separate binary logistic regression analyses were conducted to address the first research question: (a) Vocabulary and Abstraction (Composite A) as the predictors of Playthrough One Outcome and (b) Vocabulary and Block Patterns (Composite B) as the predictors of Playthrough One Outcome (see Table 6.3).

Table 6.2 Descriptive statistics and correlations

Subscales	*n*	*M*	SD	Vocabulary	Abstraction	Block patterns
Vocabulary	144	104.43	11.02	–		
Abstraction	144	107.19	15.01	0.279*	–	
Block patterns	144	107.56	14.60	0.102	0.566*	–

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

Table 6.3 Log regression analysis for cognitive ability predicting outcomes on playthrough one

Variable	*B*	SE *B*	e^B	Wald $\chi^2(1)$	*p*
Vocabulary	0.062	0.021	1.064	8.787	0.003
Abstraction	0.013	0.014	1.013	0.784	0.376
Nagelkerke R^2	0.133				
$\chi^2(2)$	13.864				0.000
Vocabulary	0.063	0.020	1.065	9.581	0.002
Block patterns	0.025	0.014	1.026	3.069	0.080
Nagelkerke R^2	0.155				
$\chi^2(2)$	16.629				0.000

Note: e^B = odds ratio

6.4.1.1 RQ1a

The logistic regression model with Vocabulary and Abstraction as predictors was statistically significant, $\chi^2(2) = 13.864$, $p < 0.001$. The model explained 13.3% (Nagelkerke R^2) of the variance in outcome and showed good fit, Hosmer-Lemeshow: $\chi^2(8) = 10.793$, $p = 0.214$. Vocabulary was a significant predictor of gameplay outcome ($B = 0.062$, SE = 0.021, $e^B = 1.064$, Wald $\chi^2(1) = 8.787$, $p = 0.003$), whereas Abstraction was nonsignificant (Wald $\chi^2(1) = 0.784$, $p = 0.376$). As participants' vocabulary scores increase, there is an associated increase in the odds of attaining a successful outcome on Playthrough One.

6.4.1.2 RQ1b

The logistic regression model with Vocabulary and Block Patterns as predictors was also statistically significant, $\chi^2(2) = 16.269$, $p < 0.001$. The model explained 15.5% (Nagelkerke R^2) of the variance in outcome and showed good fit, Hosmer-Lemeshow: $\chi^2(8) = 10.793$, $p = 0.214$. Vocabulary was again a significant predictor of gameplay outcome ($B = 0.063$, SE = 0.020, $e^B = 1.065$, Wald $\chi^2(1) = 9.581$, $p = 0.002$), while Block Patterns was nonsignificant (Wald $\chi^2(1) = 3.069$, $p = 0.080$). As participants' vocabulary scores increase, there is an associated increase in the odds of attaining a successful outcome on Playthrough One.

6.4.2 RQ2

Following the same structure as above, two separate binary logistic regression analyses were conducted to address the second research question: (a) Vocabulary and Abstraction (Composite A) as the predictors of Playthrough Two Outcome and then (b) Vocabulary and Block Patterns (Composite B) as the predictors of Playthrough Two Outcome (see Table 6.4).

Table 6.4 Log regression analysis for cognitive ability predicting outcomes on playthrough two

Variable	B	SE B	e^B	Wald χ^2 (1)	p
Vocabulary	0.027	0.016	1.027	2.666	0.103
Abstraction	0.017	0.012	1.017	1.898	0.168
Nagelkerke R^2	0.058				
$\chi^2(2)$	6.365				0.041
Vocabulary	0.030	0.016	1.030	3.384	0.066
Block patterns	0.034	0.012	1.034	7.439	0.006
Nagelkerke R^2	0.109				
$\chi^2(2)$	12.278				0.002

Note: e^B = odds ratio

6.4.2.1 RQ2a

The logistic regression model with Vocabulary and Abstraction as predictors was statistically significant, $\chi^2(2) = 6.365$, $p = 0.041$. The model explained 5.8% (Nagelkerke R^2) of the variance in outcome and showed fair fit, Hosmer-Lemeshow: $\chi^2(8) = 13.684$, $p = 0.090$. However, while the overall model was significant, neither of the individual predictors was significant (Vocabulary, Wald $\chi^2(1) = 2.666$, $p = 0.103$; Abstraction, Wald $\chi^2(1) = 1.898$, $p = 0.168$).

6.4.2.2 RQ2b

The logistic regression model with Vocabulary and Block Patterns as predictors was also statistically significant, $\chi^2(2) = 12.278$, $p < 0.002$. The model explained 10.9% (Nagelkerke R^2) of the variance in outcome and showed good fit, Hosmer-Lemeshow: $\chi^2(8) = 5.334$, $p = 0.721$. Block Patterns was a significant predictor of gameplay outcome ($B = 0.034$, SE = 0.012, $e^B = 1.034$, Wald $\chi^2(1) = 7.439$, $p = 0.006$), while Vocabulary was again nonsignificant (Wald $\chi^2(1) = 3.384$, $p = 0.066$). As participants' Block Patterns scores increase, there is an associated increase in the odds of attaining a successful outcome on Playthrough Two.

6.4.3 RQ3

To address research question three, a learning growth score was created wherein the outcome score from Playthrough One was subtracted from the outcome score from Playthrough Two. All nonpositive values were coded as a "0" for nongrowth, and all positive values were coded as a "1" for growth. Binary logistic regression models were then analyzed (Vocabulary and Abstraction as predictors of growth; Vocabulary and Block Patterns as predictors of growth). Results indicated that neither regression model was significant ($\chi^2(2) = 1.334$, $p = 0.513$; $\chi^2(2) = 3.116$, $p = 0.211$, respectively; see Table 6.5).

Table 6.5 Log regression analysis for cognitive ability predicting outcome growth

Variable	B	SE B	e^B	Wald $\chi^2(1)$	p
Vocabulary	0.008	0.016	1.008	0.228	0.633
Abstraction	0.011	0.012	1.011	0.758	0.384
Nagelkerke R^2	0.012				
$\chi^2(2)$	1.334				0.513
Vocabulary	0.009	0.016	1.009	0.341	0.559
Block patterns	0.019	0.012	1.019	2.489	0.115
Nagelkerke R^2	0.029				
$\chi^2(2)$	3.116				0.211

Note: e^B = odds ratio

6.5 Discussion

Video games are complex systems that provide contextualized experiences wherein individual differences (e.g., cognitive ability) have the potential to influence learning (Schrader et al., 2019). Despite this understanding, research into the impact of cognitive ability on performance in video games has been relatively dormant since the 1980s. The current authors set out to explicitly examine how the constructs of crystallized and fluid intelligence affect outcomes on in-game behaviors. Specifically, how do underlying cognitive mechanisms influence learning to play, and succeeding in, video games as framed by the following research questions: (1) does cognitive ability predict gameplay outcomes for the first attempt of a new game; (2) does cognitive ability predict gameplay outcomes for the second attempt of the same game; and (3) does cognitive ability predict learning growth between gameplay one and two?

Based on the results of research question one, crystallized intelligence was predictive of success for the first playthrough. As players' vocabulary scores increased, the odds of attaining success in the game on the first playthrough also increased. This may be because players relied on prior knowledge and experience (crystallized intelligence) when initially navigating the complex video game environment; the more knowledge and experience they have, the more likely they are to succeed in their initial gameplay. In addition, the first playthrough is hypothesized to act as an acclimation stage, during which players develop situated knowledge and experience, contextualized by the video game content and narrative.

This conceptualization of the first playthrough as the acclimation stage underpins the results of research question two: fluid intelligence was predictive of success for Playthrough Two. As players' scores increased, the odds of attaining success in the game on the second playthrough also increased. Over time, the reliance on, and integration of, prior extant knowledge in a novel experience diminishes. Once players acclimate to the system, they were able to better rely on problem-solving skills (i.e., fluid intelligence) to achieve successful gameplay outcomes. The presence of situated understanding was demonstrated by the statistically significant increase in success rates, from 39 to 73 successes, from the first to second playthrough ($t(143) = 4.375$, $p < 0.001$). These results align with previous findings that indicate the presence of, and need for, an acclimation stage (McCreery et al., 2011).

Finally, results of research question three indicated cognitive ability was not predictive of growth between playthroughs. The findings of research questions one and two contextualized the lack of significance associated with the third research question. Differences between playthrough performance appear to be related to acclimatization. Specifically, each playthrough experience of the game highlighted separate cognitive abilities needed to acclimate (crystallized intelligence) and progress (fluid intelligence) within the video game. As these abilities are related yet distinct ($r = 0.38$; Shipley et al., 2009), they play substantially different roles. Rather than contributing to growth across playthroughs, each of the aforementioned cognitive abilities is impacting behaviors and outcomes within their respective playthroughs. Therefore, in order to assess the impact of problem-solving within a video game, participants should possess a situated understanding of both content and narrative to draw from and then apply to the contextualized experience. The inclusion of an acclimation time period appears integral to providing participants a chance to develop this understanding.

6.6 Conclusion

Findings from the current study indicate that intra-individual variations in cognitive ability (i.e., crystalized and fluid intelligence) may differentially affect game outcomes at different stages of the game (i.e., acclimation and skill demonstration). Implications of these findings suggest that the situated understanding of context and narrative is critical when designing a research study that examines game-based outcomes. Research design should account for an acclimation stage in which respondents gain a situated understanding of the game. In the context of the current study, the lack of situated understanding in the first playthrough demonstrated the importance of intra-individual differences in both players' prior knowledge and experience. In addition to an acclimation stage, the authors further recommend that future video game study designs consider statistical controls for any notable intra-individual differences. For example, a measure of prior video game experience may strengthen researchers' ability to control for the impact of prior knowledge. Moreover, considering inter-individual differences may aid in limiting Type II errors.

The current study does come with limitations. Specifically, it sits within a sparse body of literature. As such, these findings need to be both replicated and expanded across different game contexts and genres. In addition, cognitive ability was examined in isolation of other constructs. The inclusion of crystallized and fluid intelligence within broader extant theoretical frameworks would allow for their interpretation within more comprehensive models. Further, the authors examined these constructs solely at the factor level. Future research needs to be conducted at the task level of granularity in order to examine impacts on in-game outcomes. This line of research would also work toward ameliorating the current dearth of predictive research associated with the influence of cognitive ability at the task level and video game performance.

The role of intra-individual differences is an integral component of thinking and learning. One such difference, cognitive ability, distinctly shapes how players think and acclimate within the complex video game environment. This illustrates the importance of understanding that system-based outcomes and goals are not solely the result of a designed experience; rather, they exist as the counterpart to the characteristics people bring to, and exact on, the designed experience of a video game.

References

Alexander, P., Sperl, C. T., Buehl, M. M., Fives, H., & Chiu, S. (2004). Modeling domain learning: Profiles from the field of special education. *Journal of Educational Psychology, 96*(3), 545–557. https://doi.org/10.1037/0022-0663.96.3.545

Baniqued, P. L., Lee, H., Voss, M. W., Basak, C., Cosman, J. D., DeSouza, S., Severson, J., Salthouse, T. A., & Kramer, A. F. (2013). Selling points: What cognitive abilities are tapped by casual video games? *Acta Psychologica, 142*(1), 74–86. https://doi.org/10.1016/j.actpsy.2012.11.009

Basak, C., Boot, W. R., Voss, M. W., & Kramer, A. F. (2008). Can training in a real-time strategy video game attenuate cognitive decline in older adults? *Psychology and Aging, 23*(4), 765–777. https://doi.org/10.1037/a0013494

Bavelier, D., Green, C. S., Pouget, A., & Schrater, P. (2012). Brain plasticity through the life span: Learning to learn and action video games. *Annual Review of Neuroscience, 35*(1), 391–416. https://doi.org/10.1146/annurev-neuro-060909-152832

Bediou, B., Adams, D. M., Mayer, R. E., Tipton, E., Green, C. S., & Bavelier, D. (2018). Meta-analysis of action video game impact on perceptual, attentional, and cognitive skills. *Psychological Bulletin, 144*(1), 77–110.

Beepa Pty Ltd. (2013). *Fraps [Computer software]*. Retrieved from http://www.fraps.com

Bejjanki, V. R., Zhang, R., Li, R., Pouget, A., Green, C. S., Lu, Z.-L., & Bavelier, D. (2014). Action video game play facilitates the development of better perceptual templates. *Proceedings of the National Academy of Sciences, 111*(47), 16961–16966. https://doi.org/10.1073/pnas.1417056111

Boot, W. R., Kramer, A. F., Simons, D. J., Fabiani, M., & Gratton, G. (2008). The effects of video game playing on attention, memory, and executive control. *Acta Psychologica, 129*(3), 387–398. https://doi.org/10.1016/j.actpsy.2008.09.005

Buford, C. C., & O'Leary, B. J. (2015). Assessment of fluid intelligence utilizing a computer simulated game. *International Journal of Gaming and Computer-Mediated Simulations (IJGCMS), 7*(4), 1–17.

Castaneda, L., Sidhu, M. K., Azose, J. J., & Swanson, T. (2016). Game play differences by expertise level in Dota 2, a complex multiplayer video game. *International Journal of Gaming and Computer-Mediated Simulations (IJGCMS), 8*(4), 1–24.

Cattell, R. B. (1971). *Abilities: Their structure, growth and action.* Boston: Houghton-Mifflin.

Cattell, R. B. (1987). *Intelligence: Its structure, growth and action*. Amsterdam: North-Holland.

Chase, W. G., & Ericsson, K. A. (1982). Skill and working memory. In G. W. Bower (Ed.), *Psychology of learning and motivation* (Vol. 16, pp. 1–58). Academic Press. https://doi.org/10.1016/S0079-7421(08)60546-0

Cockburn, A., Gutwin, C., Scarr, J., & Malacria, S. (2015). Supporting novice to expert transitions in user interfaces. *ACM Computing Surveys, 47*(2), 1–36. https://doi.org/10.1145/2659796

Fikkers, K. M., Piotrowski, J. T., & Valkenburg, P. M. (2019). Child's play? Assessing the bidirectional longitudinal relationship between gaming and intelligence in early childhood. *Journal of Communication, 69*(2), 124–143.

Foroughi, C. K., Serraino, C., Parasuraman, R., & Boehm-Davis, D. A. (2016). Can we create a measure of fluid intelligence using Puzzle Creator within Portal 2? *Intelligence, 56*, 58–64. https://doi.org/10.1016/j.intell.2016.02.011

Gee, E., & Gee, J. P. (2017). Games as distributed teaching and learning systems. *Teachers College Record, 119*(11), 1–22.

Glaser, R. (1985). *Thoughts on expertise*. Learning Research and Development Center.

Goh, J., Truman, B., & Barber, D. (2019). Exploring individual differences as factors to maximize interactive learning environments for future learning. *Interactive Learning Environments, 27*(4), 497–507.

Goldstein, J., Cajko, L., Oosterbroek, M., Michielsen, M., Van Houten, O., & Salverda, F. (1997). Video games and the elderly. *Social Behavior and Personality: An International Journal, 25*(4), 345–352. https://doi.org/10.2224/sbp.1997.25.4.345

Grab The Games. (2015). *The Deed [Computer software]*. Retrieved from http://store.steampowered.com/app/420740/The_Deed/

Green, C. S., & Bavelier, D. (2003). Action video game modifies visual selective attention. *Nature, 423*(6939), 534–537. https://doi.org/10.1038/nature01647

Hayes, T. L., & Thorndike, T. (2010). Review of Shipley-2. In R. A. Spies, J. F. Carlson, & K. F. Geisinger (Eds.), *The eighteenth mental measurements yearbook*. Buros Institute, University of Nebraska.

Heim, A. (1968). *The AH4 test*. National Foundation for Educational Research-Nelson.

Hilpert, J. C., & Marchand, G. C. (2018). Complex systems research in educational psychology: Aligning theory and method. *Educational Psychologist, 53*(3), 185–202. https://doi.org/10.1080/00461520.2018.1469411

Holzman, T. G., Pellegrino, J. W., & Glaser, R. (1983). Cognitive variables in series completion. *Journal of Educational Psychology, 75*(4), 603.

Jonassen, D. H., & Grabowski, B. L. (2012). *Handbook of individual differences, learning, and instruction*. Routledge.

Jones, M. B., Dunlap, W. P., & Bilodeau, I. M. D. (1986). Comparison of video game and conventional test performance. *Simulations and Games, 17*(4), 435–446. https://doi.org/10.1177/0037550086174001

Kokkinakis, A. V., Cowling, P. I., Drachen, A., & Wade, A. R. (2017). Exploring the relationship between video game expertise and fluid intelligence. *PLoS One, 12*(11). https://doi.org/10.1371/journal.pone.0186621

Kranz, M. B., Baniqued, P. L., Voss, M. W., Lee, H., & Kramer, A. F. (2017). Examining the roles of reasoning and working memory in predicting casual game performance across extended gameplay. *Frontiers in Psychology, 8*, 203.

Mané, A., & Donchin, E. (1989). The space fortress game. *Acta Psychologica, 71*(1–3), 17–22. https://doi.org/10.1016/0001-6918(89)90003-6

McCreery, M. P., Schrader, P. G., & Krach, S. K. (2011). Navigating massively multiplayer online games: Evaluating 21st century skills for learning within virtual environments. *Journal of Educational Computing Research, 44*(4), 473–493. https://doi.org/10.2190/EC.44.4.f

McCreery, M. P., Vallett, D. B., & Clark, C. (2015). Social interaction in a virtual environment: Examining socio-spatial interactivity and social presence using behavioral analytics. *Computers in Human Behavior, 51*, 203–206. https://doi.org/10.1016/j.chb.2015.04.044

McGonigal, J. (2011). *Reality is broken: Why games make us better and how they can change the world*. Penguin.

McPherson, J., & Burns, N. R. (2007). Gs invaders: Assessing a computer game-like test of processing speed. *Behavior Research Methods, 39*(4), 876–883. https://doi.org/10.3758/BF03192982

McPherson, J., & Burns, N. R. (2008). Assessing the validity of computer-game-like tests of processing speed and working memory. *Behavior Research Methods, 40*(4), 969–981. https://doi.org/10.3758/BRM.40.4.969

Minear, M., Brasher, F., Guerrero, C. B., Brasher, M., Moore, A., & Sukeena, J. (2016). A simultaneous examination of two forms of working memory training: Evidence for near transfer only. *Memory & Cognition, 44*(7), 1014–1037. https://doi.org/10.3758/s13421-016-0616-9

Miyake, A., Friedman, N. P., Rettinger, D. A., Shah, P., & Hegarty, M. (2001). How are visuospatial working memory, executive functioning, and spatial abilities related? A latent-variable analysis. *Journal of Experimental Psychology: General, 130*(4), 621–640. https://doi.org/10.1037/0096-3445.130.4.621

Murias, K., Kwok, K., Castillejo, A. G., Liu, I., & Iaria, G. (2016). The effects of video game use on performance in a virtual navigation task. *Computers in Human Behavior, 58*, 398–406. https://doi.org/10.1016/j.chb.2016.01.020

Oei, A. C., & Patterson, M. D. (2013). Enhancing cognition with video games: A multiple game training study. *PLoS One, 8*(3), e58546. https://doi.org/10.1371/journal.pone.0058546

Okagaki, L., & Frensch, P. A. (1994). Effects of video game playing on measures of spatial performance: Gender effects in late adolescence. *Journal of Applied Developmental Psychology, 15*(1), 33–58. https://doi.org/10.1016/0193-3973(94)90005-1

Quiroga, M. A., Diaz, A., Román, F. J., Privado, J., & Colom, R. (2019). Intelligence and video games: Beyond "brain-games". *Intelligence, 75*, 85–94. https://doi.org/10.1016/j.intell.2019.05.001

Quiroga, M. A., Escorial, S., Román, F. J., Morillo, D., Jarabo, A., Privado, J., Hernández, M., Gallego, B., & Colom, R. (2015). Can we reliably measure the general factor of intelligence (g) through commercial video games? Yes, we can! *Intelligence, 53*, 1–7. https://doi.org/10.1016/j.intell.2015.08.004

Rabbitt, P., Banerji, N., & Szymanski, A. (1989). Space fortress as an IQ test? Predictions of learning and of practised performance in a complex interactive video-game. *Acta Psychologica, 71*(1–3), 243–257. https://doi.org/10.1016/0001-6918(89)90011-5

Sala, G., Tatlidil, K. S., & Gobet, F. (2018). Video game training does not enhance cognitive ability: A comprehensive meta-analytic investigation. *Psychological Bulletin, 144*(2), 111–139. https://doi.org/10.1037/bul0000139

Sattler, J. M. (2008). *Assessment of children: Cognitive foundations*. JM Sattler.

Schrader, P. G., McCreery, M. P., Carroll, M. C., Head, D. L., & Laferriere, J. R. (2019). Assessing learning from, with, and in games revisited: A heuristic for emerging methods and commercial off-the-shelf games. In *Game-based assessment revisited* (pp. 13–35). Springer.

Shipley, W. C., Gruber, C. P., Martin, T. A., & Klein, A. M. (2009). *Shipley Institute of Living Scale, second edition (Shipley-2)*. Western Psychological Services (WPS).

Squire, K. (2006). From content to context: Videogames as designed experience. *Educational Researcher, 35*(8), 19–29.

Tekinbaş, K. S., & Zimmerman, E. (2004). *Rules of play: Game design fundamentals*. MIT Press.

Terlecki, M. S., Newcombe, N. S., & Little, M. (2008). Durable and generalized effects of spatial experience on mental rotation: Gender differences in growth patterns. *Applied Cognitive Psychology, 22*(7), 996–1013. https://doi.org/10.1002/acp.1420

Ventura, M., Shute, V., Wright, T. J., & Zhao, W. (2013). An investigation of the validity of the virtual spatial navigation assessment. *Frontiers in Psychology, 4*, 852. https://doi.org/10.3389/fpsyg.2013.00852

Wechsler, D. (1997). *WAIS-III administration and scoring manual*. Psychological Corporation.

Yan, E. Q., Huang, J., & Cheung, G. K. (2015). Masters of control: Behavioral patterns of simultaneous unit group manipulation in Starcraft 2. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (pp. 3711–3720) https://doi.org/10.1145/2702123.2702429

Zaval, L., Li, Y., Johnson, E. J., & Weber, E. U. (2015). Complementary contributions of fluid and crystallized intelligence to decision making across the life span. In *Aging and decision making* (pp. 149–168). Academic Press.

Chapter 7
"Study Properly": Digital Support for the Pre-Student Phase

Pia Drews and Alke Martens

7.1 Introduction

7.1.1 Junior Study Programs

Universities worldwide offer orientation programs for school students since the 1970s (e.g., Deutsche Telekom Stiftung, 2011; Solzbacher, 2008). The basic idea of these programs has been to support highly gifted school students. Historically they were launched as a side effect of allowing early performers to join university programs. In 1968, when the eighth grader Joe L. Bates participated in a summer camp at the university, the general idea was born. Mrs. D. K. Lidtke spoke to professor J. C. Stanley about the boy's outstanding talents in mathematics, informatics, and physics, so professor J. C. Stanley allowed Joe to visit university lectures. After this, the idea spread, and the first Juniorstudent class started in 1971 with 20 students from eighth and ninth grade (Stanley, 1996). This first American program is called the Study of Mathematically Precocious Youth, or SMPY. It has been shown that 8% of female school students and 18% of male school students scored equivalent or better on a college achievement test compared to regular students (Gröber, Müller, & Kuhn, 2018).

The idea of the program has been taken up and implemented differently at many universities around the world (e.g., Drews & Martens, 2019; Fatzer, 1987; Gröber et al., 2018; Solzbacher, 2011).

P. Drews (✉)
Department of Neurology, University of Rostock, Rostock, Germany
e-mail: piastine.drews@med.uni-rostock.de

A. Martens
Institute of Computer Science, University of Rostock, Rostock, Germany

D. Ifenthaler et al. (eds.), *Orchestration of Learning Environments in the Digital World*, Cognition and Exploratory Learning in the Digital Age,
https://doi.org/10.1007/978-3-030-90944-4_7

A Juniorstudy program is characterized by the fact that school students without a university entrance qualification can participate in courses and in some cases even provide certificates of achievement before the start of their studies. The legal basis is set out in the respective university laws of the states (Gröber et al., 2018).

The goal of these projects is to bind highly talented Juniorstudents to a university at an early stage. On the one hand, this is an extra qualification which supports the high level of highly gifted students and thus supports the traditional school system. On the other hand, the school students are given the opportunity of the best possible study orientation through direct experience of the university environment (Solzbacher, 2011). This should make it easier to choose a course of study and simplify the progression from school to university. Moreover, it often lifts a burden related to the parents of highly gifted students, which are not necessarily on this high intelligence level. This is in line with the main motives of the Juniorstudents, who stated that they participated in order to deepen their technical knowledge and for study and career orientation. Also to get to know university structures. However, STEM (science, technology, engineering, and mathematics) majors make up 80% of the offer. Mathematics is chosen most often, followed by physics, chemistry, and computer science (Gröber et al., 2018).

The first German early study program was founded in the winter term 2000/2001 at the University of Cologne. During the next two decades, around 50 different projects have been established in Germany. The need for an awareness of good study orientation is shown by current statistics, which show that in the last decade, consistently about 29% of students drop out of their studies during the bachelor's phase (Heublein et al., 2017).

Students who began their studies in 2014 were surveyed in a 2016 DZHW study. The results were compared with those who had started their studies in 2008. Dropout motives in 2008 and 2014 included performance problems (31 and 30%) and lack of motivation to study (18 and 17%). The financial situation of the respondents had improved significantly. For example, this was a problem for 19% in 2008, but in 2014, this issue was only mentioned by 11%. However, another important point was practical activity; this was missed by 11 and 15% in the surveyed cohorts (Heublein et al., 2017).

The dropout rate in the state examination study courses like medicine is significantly lower. On average, the dropout rate is 5–11%. However, it has been shown that 44% usually exmatriculate after six or more terms, as reported in "ÄrzteZeitung", 2017. On the other hand, there is the rush for medical school spots. For the winter term 2020/2021, 9660 medical school places were reported by the Foundation for University Admission. Three to five applications were submitted for each study place (Springermedizin, 2020).

The aim of an early study program should therefore be, on the one hand, to address reasons for dropping out and to anticipate them. On the other hand, it could also be a good basis for selecting the right applicants for a small number of study places (e.g., for the study places for human medicine).

With these ideas, the junior study program of the University of Rostock was founded in the winter term 2008/2009.

7.1.2 Online vs. Presence

In 2018, Gröber et al. showed that there were about 50 different early learning programs in Germany at that time. Most of them were presence-based offerings. While these can provide an unfiltered look at a lecture course, they can usually only be accessed by school students from a narrow radius around the university itself. In addition, school lessons are missed in favor of lectures, which must be made up independently.

Only three universities offer online or blended learning-based early learning: Distance Learning University of Hagen, Technical University of Kaiserslautern, and the University of Rostock (Gröber et al., 2018).

On the website of the Technical University of Kaiserslautern, it can be read that a social contribution of 245.23€ (as of winter term 2020/2021) has to be paid. In addition, only mathematics and physics are available as an online early study program.

According to information from the website of the Distance Learning University of Hagen, the selection of the lecture courses is much larger. Nevertheless, here, too, the regular study fee must be paid per module.

Both universities include Juniorstudents in their regular distance learning programs. According to our information, there is no targeted processing of the learning material for Juniorstudents.

7.1.3 Juniorstudy Program of the University of Rostock

The junior study program at the University of Rostock (if the junior study program of the University of Rostock is meant, it will be called Juniorstudy in the following) organizes things differently: When the project was founded in the winter term 2008/2009, lectures from the basic studies (usually first–third term) were recorded. These are offered free of charge to all enrolled Juniorstudents on a learning platform (Stud.IP) at weekly intervals. The Juniorstudy is designed for students in 10th to 12th grade. At the end of the 12th grade, they achieve the baccalaureate. The preparatory phase is called Sekundarstufe II—Sek II.

The selection process is simple and does not require confirmation of giftedness from the school. The Juniorstudy is not dependent on any local ties—school students from all over Germany (and other German-speaking countries) can participate. In addition, the project is open to all who are interested in study orientation. In this way, it can be offered to a comprehensive study orientation and preparation to a large number of young people.

As indicated by the increasing number of participants (see Fig. 7.1), interest in junior studies has been steadily increasing over the past 11 years.

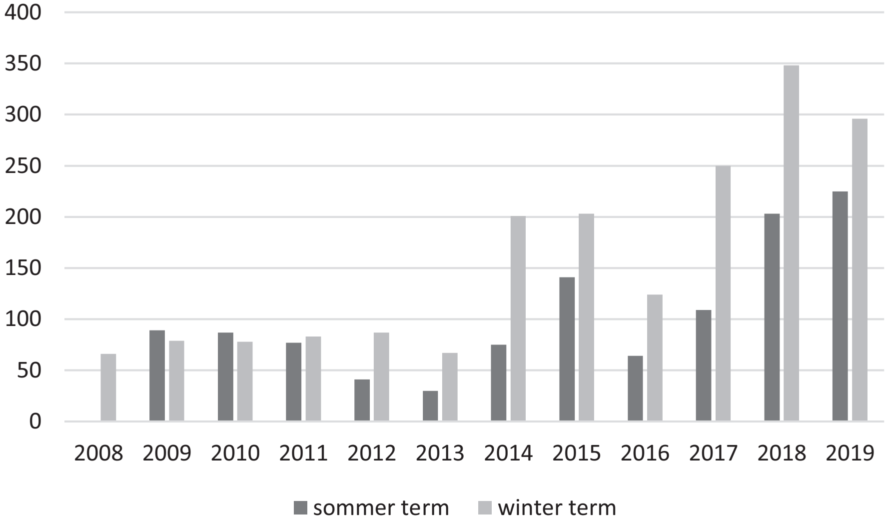


Fig. 7.1 Increase of the participating Juniorstudents since winter term 2008/2009

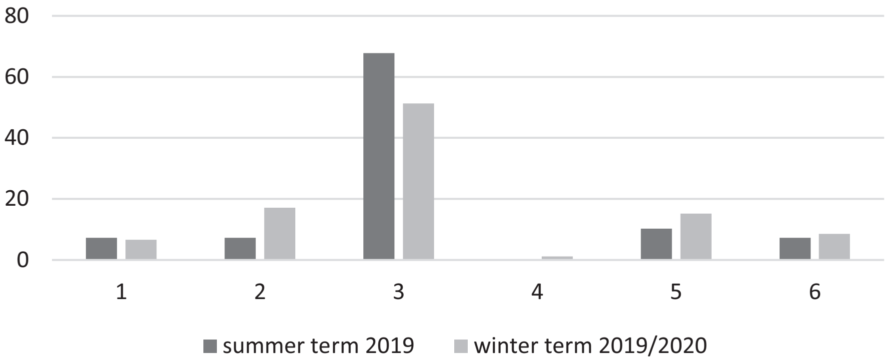


Fig. 7.2 Distribution of the chosen courses in summer term 2019 and winter term 2019/2020. *1* Engineering and computer sciences, *2* mathematics and natural sciences, *3* medicine/life sciences, *4* linguistics and humanities, *5* economics, social sciences, and law, and *6* teaching professions

Currently the Juniorstudy offers around 20 different lecture courses including study courses such as medicine, STEM fields, or teaching. Figure 7.2 shows the distribution of subject choices in summer term 2019 and winter term 2019/2020.

Each lecture course is accompanied by a student tutor who can answer questions regarding content and organization and thus bridge gaps in knowledge. In addition, they also share their subjective study experience. This gives the Juniorstudents a comprehensive insight into the everyday life of the university.

During the term, Juniorstudents get a chance to practice what they have learned through exercises and tests. In addition, the Juniorstudents are invited to Rostock at least twice per term to participate in optional seminars in order to apply the theoretically acquired knowledge. This should give an even better impression of everyday student life. At the end of the term, a final exam may be taken in a few subjects at the end of the term. If you pass this, the certificate can be accredited in the form of credit points in their later studies.

To our knowledge, the Juniorstudy is unique in the German-speaking world. It is therefore even more interesting that the learning performance of Juniorstudents was analyzed for the first time in this study using the example of “Chemistry for Medicine.”

7.1.4 Learning Platform “Stud.IP”

The online learning platform used is Stud.IP, which is also offered to regular students at the University of Rostock.

This section will introduce the learning platform and its features.

As shown in Fig. 7.3, a Juniorstudent can see the chosen course on the home page of Stud.IP. With one the Juniorstudent gets access to the course and all the tools that are used during the term.

Essential features of the learning platform functions are shown in Fig. 7.4. The home page of a course is shown. Tutors can use a survey option, for example, to make appointments with Juniorstudents.

Both the tutor and Juniorstudents use the forum to share information and ask questions. At the same time, tutors can initiate group work.

Information about the term schedule and tasks will be posted in the data folder. Another tool is the so called Wiki. It provides Juniorstudents with the opportunity to compile their lecture summaries. The tutor reads these summaries several times during the term to correct errors. This enables Juniorstudents to prepare for

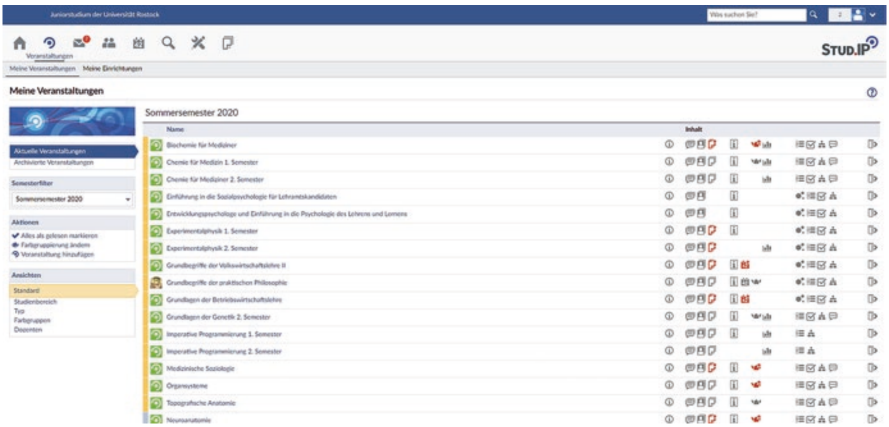


Fig. 7.3 Home page of Stud.IP

Fig. 7.4 An overview of the various tools used in the event is shown. *1* forum, *2* data, *3* Wiki, *4* tasks, *5* vips, *6* grades, *7* lecture videos

assignments and tests together. These tests can be created very differently, thanks to the features available on the platform. On the one hand, there is the possibility to provide open questions and worksheets. On the other hand, tests with multiple-choice or single-choice tasks can be prepared. However, these are only frequently used functions; depending on the module and regular study program, other test formats may also be used. After the submitted solutions are corrected, Juniorstudents receive individual feedback. The score achieved is additionally entered in the grades list. Finally, there is a tool in the course to provide the lectures. Here, the tutors upload a lecture every week, which can then be viewed flexibly in terms of time and location during the course of the term.

7.2 Learning Success of Juniorstudents

Online teaching and blended learning formats as part of the study course have already been evaluated frequently. A study was also conducted on this at the University of Rostock. Since the junior study program also recorded lectures from the field of "social psychology," these videos could be used to compare an online course with the same content as a regular face-to-face course (visited by regular students, not by Juniorstudents). The results showed that performance was slightly better in the online offering and that the grades "very good" and "good" were achieved more frequently (Perleth & Neumann, 2011).

E-learning courses in medical studies were also analyzed at the University of Erlangen and compared with face-to-face courses for students. The results showed that the learning success was comparable. The online teaching program was better evaluated. However, the classical teaching system was subjectively preferred (Triebner, 2010).

As far as literature review by the authors has shown, there has been little evaluation of Juniorstudents’ learning success. Therefore, this article is to confirm the previous study by Drews and Martens (2020), which showed a great learning success of the Juniorstudents. But before the methods and results are to be described, the Juniorstudents and their diverse backgrounds are described below.

7.2.1 Interesting Facts about the Juniorstudents

The Juniorstudy is evaluated for quality assurance every term. This evaluation includes how many lecture videos were actually viewed, whether the term performances were manageable in addition to school, and whether the school students have ideas for further development of the Juniorstudy.

A few terms ago, a closer look at the background of the Juniorstudents was started to find out how to optimize the project further. On the other hand, the aim was to find out if and how Juniorstudents could be given more ideas for learning and study strategies. Since this project does not specialize only in high-ability school students, the composition of Juniorstudents is diverse, and the challenge is to accommodate everyone.

This study surveyed Juniorstudents in the summer term 2019 and winter term 2019/2020. Therefore, the following information refers to these two terms. General information about all Juniorstudents is presented first, followed by facts about Juniorstudents interested in medicine.

A total of 417 Juniorstudents participated in both terms. 67.3% of them were female. 68.6% stated their place of residence in Mecklenburg-Western Pomerania, 50.7% of those directly in Rostock.

On average, participants were 16.7 years old.

19.2% stated that they had already successfully completed school. The distribution of school classes can be seen in Fig. 7.5.

This means that 71.7% were in the specifically set target group (10th–13th grade).

One hundred twenty-two of 417 Juniorstudents (29.3%) participated for at least one term as part of their school’s elective classes. Thirty Juniorstudents even decided to take more than one course. 50.4% completed the Juniorstudy after one term. Two hundred seven Juniorstudents participated in two to seven terms, as shown in Fig. 7.6 (last checked 15.02.2021).

Of these 417 participants, 251 Juniorstudents (60.2%) attended a medical course. Significantly more—74.9%—were female. 60.2% lived in Mecklenburg-Western Pomerania, of which 41.7% stated their residence directly in Rostock. More participants of the medical field attended with higher distance to the event location Rostock.

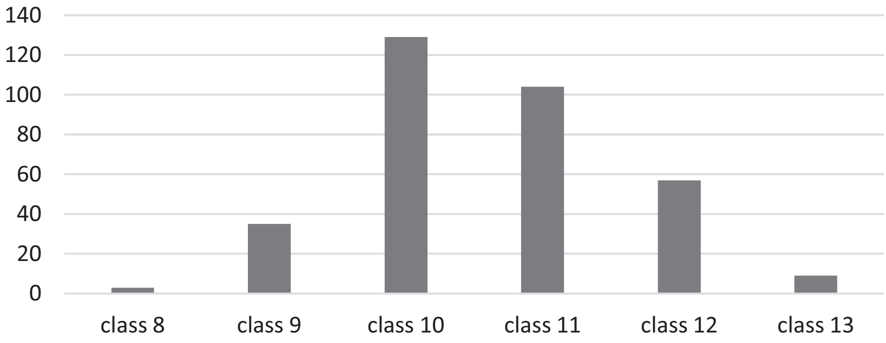


Fig. 7.5 Number of participants of different school classes

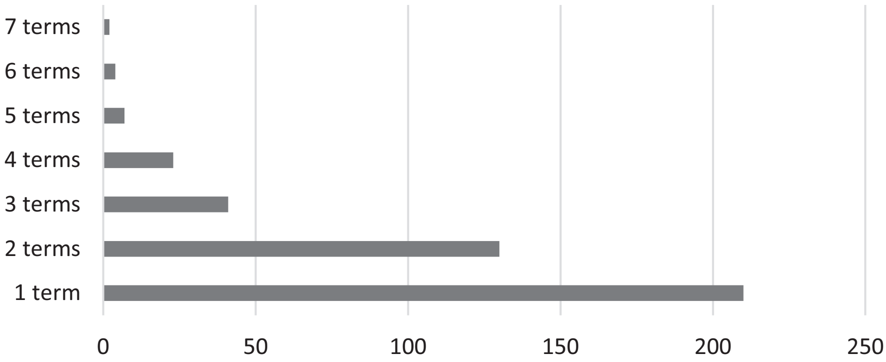


Fig. 7.6 Number of participants, sorted according to the number of terms attended

Medical Juniorstudents are also older on average (17.6 years). This can be explained by the fact that especially the anatomical modules and "Chemistry for Physicians" have an age limit of at least 16 years.

29.1% of the medical Juniorstudents stated that they had already finished school. In absolute terms, this corresponds to 73 participants. Thus, in the period studied, those who had already finished school almost exclusively chose a medical module. This is particularly interesting because many people interested in medicine in Germany do not get a place immediately after graduating from school. Early study programs seem to be an option to bridge the waiting period and to provide further education.

Only 10% of the medical Juniorstudents took part in the Juniorstudy as an elective subject of their schools. On the one hand, this is certainly due to the fact that a large proportion no longer went to school; on the other hand, most compulsory elective school students are in the ninth or tenth grade and accordingly do not yet reach the age limit for some medical subjects.

The number of courses attended by medical Juniorstudents is comparable to the overall average for all Juniorstudents: 49.8% attended one term. However, 73

Juniorstudents (29.1%) completed 2 terms, and another 21.2% participated in 3 to a maximum of 7 terms (last checked 15.02.2021).

It is difficult to make an exact statement about the dropout rate. Since the Juniorstudy of the University of Rostock is offered to a broad group of school students, everyone is free to participate in the term exercises and lectures. Thus, only the statement can be made about those students who regularly participate in the term offerings. However, due to data protection reasons, it cannot be evaluated whether someone has dropped out of the Juniorstudy or only attended the lectures.

All Juniorstudents who participated in the first face-to-face course in summer term 2019 and winter term 2019/2020 were invited to fill out paper-based questionnaire to get more background information about the school students. In addition, school students not participating in the Juniorstudy of the University of Rostock were also asked to complete the questionnaire as an online version to have a comparison group. This online questionnaire was also answered by 20 Juniorstudents during the summer term of 2020. These were also included in the following data summary.

Of the total 164 questionnaires evaluated, 118 had participated in a medical module. The average age of the respondents was 17.1 years. Female participants predominated with 75%. Fifty-five of the Juniorstudents (n = 52 medical, n = 3 nonmedical) stated their current grade average or final school grade. In Germany, grades are assigned from 1 to 6, with 1 corresponding to very good and 6 to fail. On average, this grade was 1.6, with a minimum of 1.0 and a maximum of 2.7.

The detailed analysis of the results of the questionnaire will not be discussed here. However, it is interesting to note that 58.4% of the mothers (n = 137) and 58.9% of the fathers (n = 134) had a university degree. Among medical Juniorstudents, 57.6% of mothers and 63.3% of fathers were reported to have a college degree (both n = 97). In 71.7%, at least one parent had a university degree. This is also consistent with Solzbacher's, 2011 data collection, which found that 71.3% of the school students she had questioned had at least one parent with an academic background. In 40.2%, at least one parent worked in the medical sector.

7.2.2 Study Background and Previous Results

As shown above, the Juniorstudy at the University of Rostock is characterized by a unique offer for prospective students. However, even if the offer is subjectively perceived as a great enrichment for the students, the question arises whether there is also an added learning value.

This information is potentially important because the school system in Germany allows school students to choose their subjects very diversely, especially in the last years before they graduate from high school. On the other hand, however, it is hardly possible for a school student to continue taking all the natural sciences after the tenth grade. Thus, the different choice of school subjects already leads to different learning levels at the beginning of studies. However, in the study of human medicine, for example, one needs the basics of all natural sciences in order to be able to understand further subjects such as "biochemistry" or "physiology." Thus, the challenge of the introductory study phase is to bring all students to a level of prior knowledge in the shortest possible time.

Since study places are sometimes allocated at very short time before the start of studies, there is no time left to offer intensive pre-study courses. This raises the question of whether, for example, an early study program can not only provide support in the selection of a course of study, but can even make the introductory phase of a course of study substantially easier by teaching the content of the course.

For this reason, the subjective and measurable learning success was evaluated for the first time in the summer term 2019 and winter term 2019/2020. For this purpose, the module "Chemistry for Medicine" was selected as an example, since it is the best-attended course.

The teaching content of "Chemistry of Medicine" includes a total of 21 lectures (each 105 min) and 7 practicals (90-min experiment demonstration and 90-min training theoretical tasks) divided over 2 terms for the Juniorstudents (in the regular course of study, regular students must manage this course content in 1 term). Juniorstudents who participated in this study attended ten lectures and four practicals.

The course content of "Chemistry for Medicine" in the first of two terms is mostly a repetition of the chemistry lessons of the school with a focus on general chemistry. Not shown in this study is the second term, which deals more with organic chemistry.

As already shown in Drews and Martens (2020), the Juniorstudents had major problems with mathematical basics. These were simple questions such as "switch to X" or the "conversion of units." At the beginning of the term, it had been communicated that such basics would be part of the final exam, but would not be explicitly taught in the course.

However, the learning success in the area of chemistry mathematics was astonishing. Here, the Juniorstudents achieved significantly better results in five out of six tasks at the end of the term.

In the following, the learning success will be better illuminated on the basis of further tasks, and, in addition, the subjective assessment of the Juniorstudents will be considered.

7.3 Methods

In this study, Juniorstudents from the lecture course "Chemistry for Medicine" were surveyed. Only the first-term participants were included. The first term focuses on general chemistry. Due to the small group size in the second-term "Chemistry for Medicine," this was not initially evaluated.

The survey was conducted in the summer term 2019 and the winter term 2019/2020. Two hundred ten Juniorstudents were enrolled in summer term 2019. Of these, 26 have chosen to enroll in “Chemistry for Medicine.” In winter term 2019/2020, 51 out of 296 enrolled Juniorstudents chose “Chemistry for Medicine.”

Out of a total of 77 enrolled “Chemistry for Medicine” Juniorstudents, 23 participated in the study. This corresponds to a response rate of 29.9%.

In order to assess the knowledge level of the Juniorstudents, they had to participate in a pretest at the beginning of the term. This lasted 90 min. It included a 10-min section on mathematical basics, that had to be answered without pocket calculator or other tools. Furthermore, they had to answer different tasks about general chemistry in 80 min. During the second part, a nonprogrammable calculator, the periodic table of the elements, and a small collection of chemical formulas were allowed. It consisted of 35 tasks in total.

Following this, the curriculum consisting of lectures, practicals, and exercises was completed. Finally, the posttest had to be taken after a 3-month term. At the beginning of the term, it was announced that the posttest would be close to the pretest. To ensure comparability, the same tasks were used at the end of term—contrary to the information the Juniorstudents had received.

Both—pre- and posttest—were preceded by a questionnaire on the subjective self-assessment of one’s own chemical knowledge. The questionnaire comprised 27 questions about different chemical subjects like mathematical skills, nomenclature, or redox reaction. In this study, nine questions related to the tasks shown in Fig. 7.7 will be discussed. The subjective self-assessment was rated in a 6-point scale ranging from 1 (very confident) to 6 (not confident at all). Twenty-two Juniorstudents filled in the self-assessment.

As a comparison group, regular students of the first term were asked to complete the tasks as well. Twenty-two regular students participated in winter term 2019/2020 at the beginning of their studies. Thus, they had not yet completed the lecture course “Chemistry for Medicine” and only had to answer with their prior knowledge from their school days.

The tasks shown in Fig. 7.8 (translated from German into English) are analyzed in this study.

SPSS (Version 27, IBM Corp., 2020) was used for statistical analysis. First, it was verified that the data did not have a normal distribution. Pre- and posttest of the Juniorstudents were evaluated with the Mann-Whitney U test. For the comparison of the posttest of the Juniorstudents with the regular students, the Wilcoxon test was used.

7.4 Evaluating the Results

Twenty-three of 77 Juniorstudents had participated in the study during summer term 2019 and winter term 2019/2020. The age distribution of Juniorstudents and regular students is shown in Fig. 7.9. The mean age of the Juniorstudents was 17.9 years, and the mean age of the regular students was 20.1 years.

7. Write down which states of aggregation you know.

8. Explain what sublimation is.

9. State the factors on which the melting point of ice depends.

10. Name the following salts.

a) $BaSO_4$ *b)* $AlCl_3$ *c)* Na_cCO_3

11. Give the chemical formulas for the following compounds.
a) sulfuric acid b) acetic acid c) calcium chloride

12. Name the types of radioactive radiation and characterize them.

13. The samarium isotope is an alpha-emitter. Indicate which element is formed during radioactive decay.

14. The gold isotope is a beta-(-)-emitter. Indicate which element is formed during radioactive decay.

15. The potassium isotope is a beta-(-)-emitter. Indicate which element is formed during radioactive decay.

18. Explain how the elements are ordered on the periodic table.

19. Explain what information the periodic table provides about the properties of the elements.

20. Oxygen has the atomic number 8. What can you deduce from it?

27. Draw an energy scheme for an uncatalyzed and a catalyzed reaction.

32. Set up the redox equation for the conversion of $KMnO_4$ *with* $NaSO_3$ *in acidic solution.*

Fig. 7.7 Excerpt from the test. All tasks examined in this study are shown

4. I believe that my knowledge of the manifestations of matter and their structure is very good.

5. I am good at handling the nomenclature of anorganic substances.

6. I can work well with definitions and calculations in the field of radioactive radiation.

8. I feel confident with the electron configuration.

9. I feel confident working with the periodic table of the elements.

10. I know well about chemical bonds.

11. I am familiar with intermolecular interactions.

15. I feel safe in the kinetic consideration of chemical reactions

23. I feel confident in setting up redox reactions.

Fig. 7.8 Extract of the questionnaire for subjective assessment

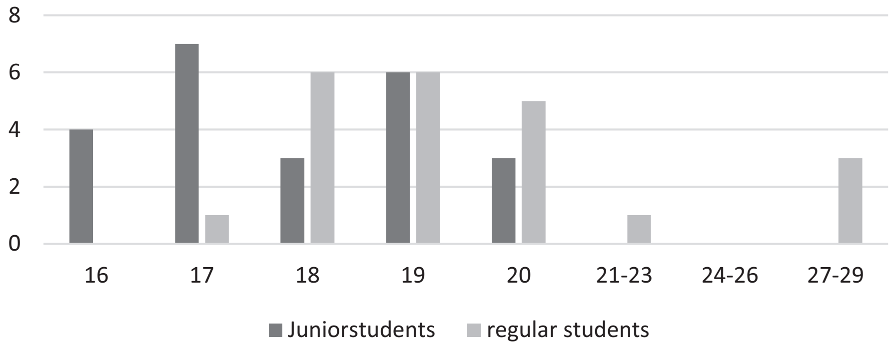


Fig. 7.9 Age of the participants. Mean age Juniorstudents 17.9. Mean age regular students 20.1

As already shown in Drews and Martens (2020), 6 of 23 Juniorstudents would have passed the test already in the beginning of the term (with a passing limit of 50%). At the end of the term, 19 of 23 Juniorstudents passed the test. They reached a mean total score of 66%. Compared to this, it has already been shown that of the regular students, only 1 out of 22 would have barely passed the test!

Regarding the group of investigation, it can be seen that some of the students are over 20 years old. These students are waiting on a place to study medicine in Germany. This waiting time can take up to 7 years. Thus, when those students finally get a place at the university to enter the first term, they have often finished another qualification—often a medical one (e.g., caregiver). However, for the study, they are still exemplary for those classical first-term students, which did not get a place at the university directly after finishing school. Thus we have integrated them in the study as well.

In the set of first-term students, three persons had finished school with a baccalaureate in 2017, five persons finished in 2018, and nine persons in 2019. From two persons, the information is missing. Some have older baccalaureates as they had waited for a place for study medicine. Regarding the control group, the baccalaureate is not that far away to assume a complete loss of chemistry knowledge.

As already shown in Drews and Martens (2020), an additional criterion for distinction has been the chemistry course at Sek II, or even chemistry as selected course in their baccalaureate. Regarding the control group, only 2 persons have had chemistry as part of their baccalaureate exams (i.e., intensive course), and 13 persons have had chemistry in their Sek II. Seven of 23 Juniorstudents already passed their baccalaureate, 7 Juniorstudents are in 11th grade, and 8 Juniorstudents are in their final school year.

The test was taken by both groups, the Juniorstudy group and the new enrolled students of medicine. The Juniorstudents were questioned in summer term 2019 and winter term 2019/2020 at both the beginning and end of the term. The students were questioned in winter term 2019/2020 in the beginning of their studies.

In the posttest, the Juniorstudents succeeded in 24 of 35 tasks and reached 66% of the total. Taking the 50% benchmark for passing the exam, 19 of the 23

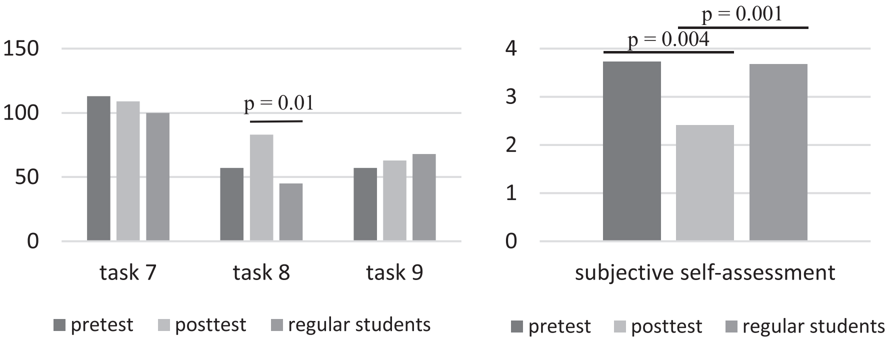


Fig. 7.10 Results of tasks 7, 8, and 9 as well as the subjective self-assessment of Juniorstudents in the pre- and posttest and of the regular students

Juniorstudents would have passed the real exam. This is incredible, compared with the 1 (one!) person passing in the enrolled student group, mentioned above (there, one person out of $n = 22$ has passed the exam).

Based on this information, which has already been published in Drews and Martens (2019), further questions of the questionnaire will now be evaluated. The chemistry test was completed by 23 Juniorstudents and 22 regular students. The subjective assessment questionnaire was completed by 22 Juniorstudents and 22 regular students.

Tasks 7, 8, and 9 (see Fig. 7.10) address the characteristics of matter. Task 7 was answered with full score by the Juniorstudents in both the pretest and the posttest. Additional points could even be achieved. Tasks 8 and 9 were answered better on average at the end of the term, but the improvement was not significant.

Subjectively, however, the Juniorstudents had the impression that they had learned more in this area. On a 6-point scale (1 = very confident to 6 = not confident at all), they were unsure at the beginning of the term ($mean_{pre\text{-}test} = 3.73$) and felt significantly more confident at the end of the term ($mean_{post\text{-}test} = 2.41$; $p = 0.004$).

Twenty-one of the 22 first-term students also achieved 100% correct scores on task 7. They performed significantly worse in task 8: While the Juniorstudents achieved around 83% of the points at the end of the term, the regular students only managed 45% on average. This showed a significantly better teaching performance of the Juniorstudents ($p = 0.01$). This also reflects the subjective assessment of both groups ($mean_{Juniorstudents} = 2.41$, $mean_{regular\ students} = 3.68$, $p = 0.001$).

Tasks 10 and 11 (see Fig. 7.11) referred to chemical nomenclature. The Juniorstudents already achieved a good result in question 10 at the beginning of the term (mean = 84%), so that no significant improvement was possible here. At the end of the term, all Juniorstudents answered the question 100% correctly! In question 11, they improved from an average of 62% to an average of 96% (2 of 23 did not achieve the full score). This corresponds to a significantly better result ($p = 0.003$).

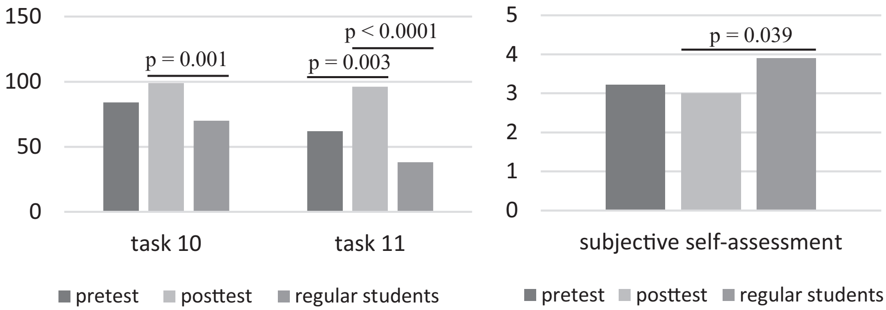


Fig. 7.11 Results of tasks 10 and 11 as well as the subjective self-assessment of Juniorstudents in the pre- and posttest and of the regular students

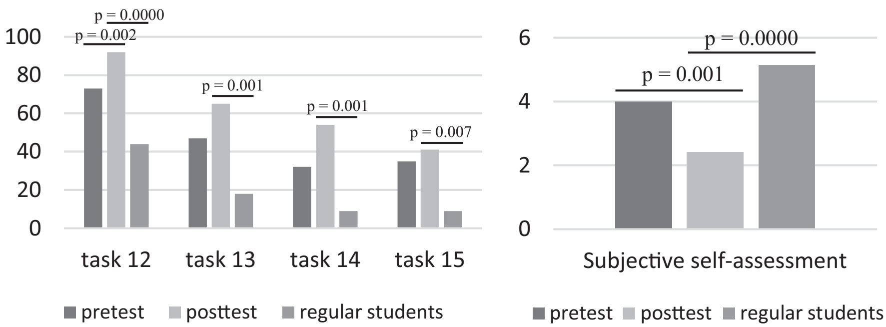


Fig. 7.12 Results of tasks 12–15 as well as the subjective self-assessment of Juniorstudents in the pre- and posttest and of the regular students

Regular students scored an average of 70% (task 10) and 38% (task 11), performing significantly worse than Juniorstudents ($p_{task10} = 0.001$, $p_{task11} < 0.0001$).

Subjectively, Juniorstudents felt insecure both at the beginning and at the end of the term ($mean_{pre\text{-}test} = 3.23$, $mean_{post\text{-}test} = 3.00$). However, they rated themselves significantly better than the regular students ($mean_{regular\ students} = 3.90$; $p = 0.039$).

Tasks 12–15 related to theoretical principles of radiation and practical skills. The Juniorstudents had understood the theoretical basics significantly better at the end of the term ($mean_{pre\text{-}test} = 73\%$, $mean_{post\text{-}test} = 92\%$; $p = 0.002$). Application, on the other hand, still seems to be a major problem, as can be seen in Fig. 7.12.

It was interesting to note that Juniorstudents rated themselves significantly better in the theoretical principles of radiation at the end of the term ($mean_{pre\text{-}test} = 4.00$, $mean_{post\text{-}test} = 2.41$; $p = 0.001$).

Comparing the results of the Juniorstudents with those of the regular students, it was surprising to find that the regular students did not pass any of the tasks. In the basics, they scored an average of 44%. They were significantly worse than the Juniorstudents ($p < 0.0001$). In tasks 13–15, only 2–4 regular students had approaches to solutions.

Task 16 (see Fig. 7.13) referred to the knowledge of electron configurations. Here, the Juniorstudents achieved only an insufficient learning success ($mean_{pre\text{-}test}$ = 41%, $mean_{post\text{-}test}$ = 65%; $p = 0.062$). Among the regular students, only three students had an approach to the task ($mean_{regular\ students}$ = 9%). Thus, even though the overall result is not yet convincing, the Juniorstudents were significantly better than the regular students ($p < 0.0001$).

Subjectively, Juniorstudents rated themselves significantly better at the end of the term ($mean_{pre\text{-}test}$ = 3.27, $mean_{post\text{-}test}$ = 2.27; $p = 0.023$). Regular students had a realistic opinion of their own knowledge ($mean_{regular\ students}$ = 4.50).

Tasks 18, 19, and 20 (see Fig. 7.14) addressed the periodic table of the elements as well as properties and information that can be deduced from it.

Question 18 was passed after average at the end of the term ($mean_{pre\text{-}test}$ = 35%, $mean_{post\text{-}test}$ = 65%; $p = 0.108$), but 8 of 23 Juniorstudents had no approach to answer the question at the end of the term.

The learning success in task 19, in which the goal was to name the essential information that can be concluded from the periodic table of the elements, was pleasing. Whereas the Juniorstudents had hardly any knowledge about this at the beginning of the term, by the end of the term 13 out of 23 Juniorstudents had correctly completed the task ($mean_{pre\text{-}test}$ = 17%, $mean_{post\text{-}test}$ = 57%; $p = 0.004$).

Task 20, in turn, was intended as an example task of the theory knowledge queried and was already well answered by the Juniorstudents at the beginning of the term ($mean_{pre\text{-}test}$ = 85%, $mean_{post\text{-}test}$ = 98%; $p = 0.084$).

The regular students performed similarly to the Juniorstudents: Tasks 18 and 19 were difficult for the students, but they were able to answer the practical application in task 20 surprisingly well ($mean_{task18}$ = 45%, $mean_{task19}$ = 28%, $mean_{task20}$ = 82%).

Subjectively, both groups rate themselves similarly ($mean_{pre\text{-}test}$ = 2.82, $mean_{post\text{-}test}$ = 2.09, $mean_{regular\ students}$ = 2.59).

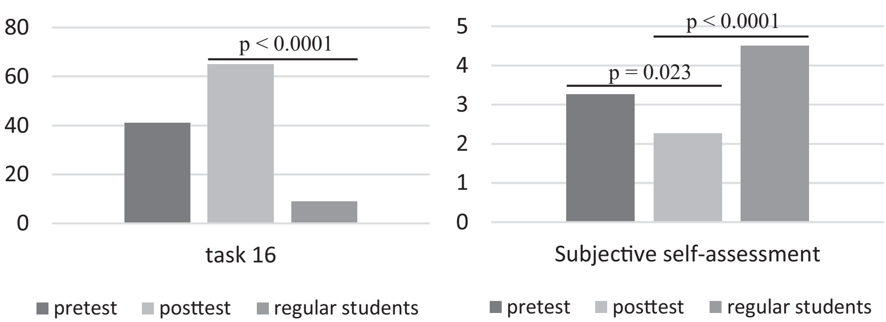


Fig. 7.13 Results of task 16 as well as the subjective self-assessment of Juniorstudents in the pre- and posttest and of the regular students

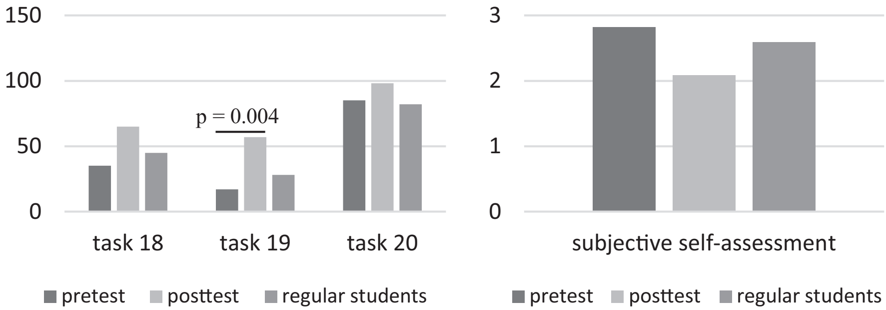


Fig. 7.14 Results of tasks 18–20 as well as the subjective self-assessment of Juniorstudents in the pre- and posttest and of the regular students

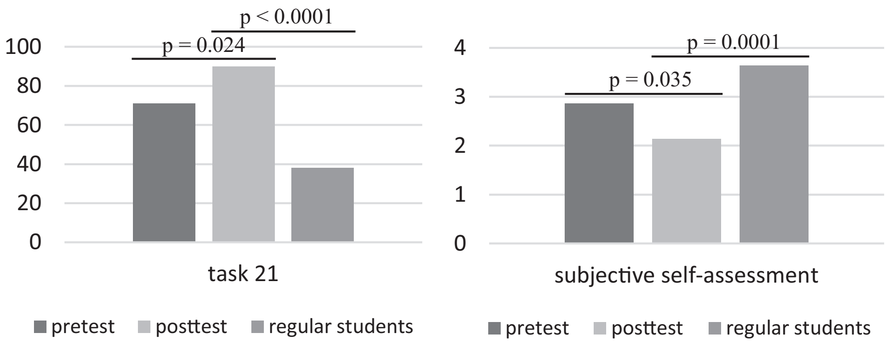


Fig. 7.15 Results of task 21 as well as the subjective self-assessment of Juniorstudents in the pre- and posttest and of the regular students

On the other hand, in task 21, respondents were asked to consider types of bonding (see Fig. 7.15). Here, the Juniorstudents improved significantly ($mean_{pre\text{-}test} = 71\%$, $mean_{post\text{-}test} = 90\%$; $p = 0.024$). The regular students found this question very difficult ($mean_{regular\ students} = 38\%$), so they performed significantly worse ($p < 0.0001$).

The subjective assessment was also in line with this learning performance ($mean_{pre\text{-}test} = 2.86$, $mean_{post\text{-}test} = 2.13$; $p = 0.035$). The regular students also assessed themselves as matching their performance ($mean_{regular\ students} = 3.64$).

Task 27 was about catalysts (see Fig. 7.16). Juniorstudents found this task particularly difficult at the beginning of the term. Even by the end of the term, 12 of 23 Juniorstudents would not have passed the task ($mean_{pre\text{-}test} = 4\%$, $mean_{post\text{-}test} = 54\%$; $p < 0.0001$).

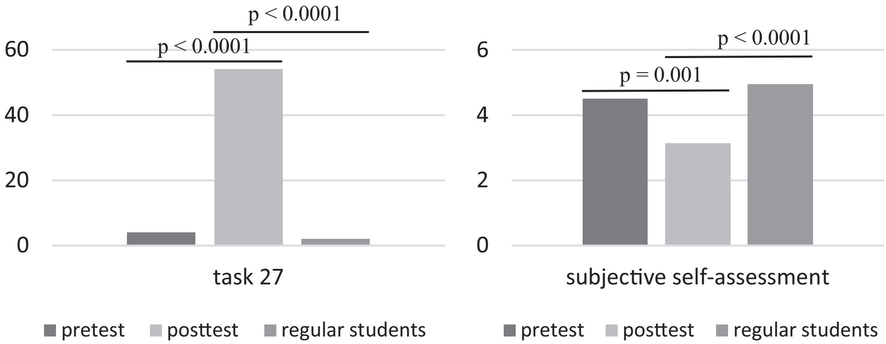


Fig. 7.16 Results of task 27 as well as the subjective self-assessment of Juniorstudents in the pre- and posttest and of the regular students

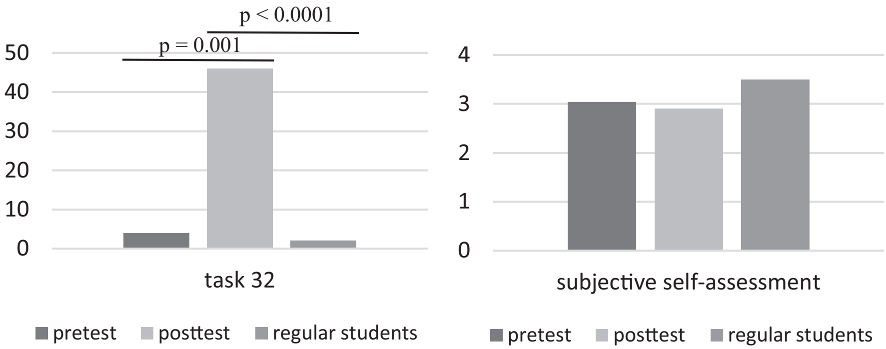


Fig. 7.17 Results of task 32 as well as the subjective self-assessment of Juniorstudents in the pre- and posttest and of the regular students

Two of the regular students had at least some approach to completing the task. None of the 22 participants would have passed the question ($\text{mean}_{\text{regular students}} = 5\%$).

Juniorstudents also rate themselves well on this task ($\text{mean}_{\text{pre-test}} = 4.50$, $\text{mean}_{\text{post-test}} = 3.14$; $p = 0.001$). The regular students rate themselves very poorly ($\text{mean}_{\text{regular students}} = 4.95$).

Lastly, task 32 was to solve a redox reaction. Only one Juniorstudent was able to answer the question correctly at the beginning of the term. All other 22 did not have a clue how to set up a redox reaction. The situation was different at the end of the term: 11 of 23 Juniorstudents had passed the task ($\text{mean}_{\text{pre-test}} = 4\%$, $\text{mean}_{\text{post-test}} = 46\%$; $p = 0.001$).

The task was also very difficult for the regular students. No one was able to solve the task ($\text{mean}_{\text{regular students}} = 0\%$).

Although this test score was poor, both regular students and Juniorstudents rate themselves as average ($\text{mean}_{\text{pre-test}} = 3.05$, $\text{mean}_{\text{post-test}} = 2.91$, $\text{mean}_{\text{regular students}} = 3.50$) (Fig. 7.17).

7.5 Conclusion and Outlook

This study is intended to complement our previous results from Drews and Martens (2020). Most recently, we had been able to acquire 23 Juniorstudents from the field of “Chemistry for Medicine” for a study, who had completed a pretest and a posttest in order to record learning success in the Juniorstudy. In the previous study, 12 of the 35 questions on mathematics and chemistry mathematics had been evaluated first. This had shown that the Juniorstudents had gained a tremendous amount of knowledge and had performed significantly better by the end of the Juniorstudy term than regular students at the beginning of their first term.

As we were able to show in the previous section, the analysis of the other questions supports the thesis that Juniorstudents have immense learning success and thus perform better overall than first-term students.

It was shown that the Juniorstudents had already performed very well in three tasks at the beginning of the term and thus hardly any further improvement was possible. In turn, significant improvement was achieved in six tasks. However, knowledge gaps were also revealed, for example, topics such as dealing with practical tasks on radioactive radiation, bonding types, catalysts, and redox reactions remain weak points for the Juniorstudents. With this knowledge, teaching can be adapted in future terms of the Juniorstudy.

It was possible to deduce from the data that the Juniorstudents are overall good to very good at self-assessment. This means that the subjective opinion of the Juniorstudents can be relied upon in part when planning the courses and the focus can be adjusted accordingly.

Subsequently, this study should be repeated, and thus the newly implemented teaching concept should be evaluated.

The results of the regular first-term students are alarming. They scored significantly lower than the Juniorstudents in 11 out of 16 questions. On average, they would have passed only seven tasks (with a pass rate of 50%).

As described earlier, most early study programs are used to provide students with extensive study orientation on the one hand and to tie them to a university at an early stage on the other hand.

However, the results of this study also suggest that the learning success in early study programs can be immense and thus it could be an option to close the gap between the acquired school knowledge and the necessary prerequisite knowledge for a study course.

We would like to support that junior studies is an ideal chance to bring together committed and interested students with different learning levels and to equalize the learning level. It thus offers excellent preparation for university studies and can facilitate the completion of undergraduate courses.

Thanks to the concept of a blended learning event, participation is not linked to physical proximity to the university. In addition, the low-threshold application procedure makes participation possible for anyone who is interested.

The long-term success of the Juniorstudents cannot yet be determined. Further studies are planned in this regard. Among other things, the Juniorstudents are to be asked retrospectively whether the Juniorstudy program has helped them to cope with the introductory phase of their studies. This should not only focus on objective results such as academic success but also on subjective perceptions such as finding one's way in the daily routine of studying.

In addition, this type of study must be transferred to other modules in order to check whether the learning success can be replicated in other modules.

Are you interested in starting a similar project at your university? We are at your disposal to initiate the first steps.

Acknowledgments This acknowledgment is for Dr. Gisela Boeck, our brilliant lecturer of "Chemistry for Medicine" and one of our greatest supporters of the Juniorstudy. Another thank goes to the student tutors, who have and are working in our Juniorstudy. And last but not least, we thank our University of Rostock, the financier of this project.

References

ÄrzteZeitung. (2017). *Warum Medizinstudenten ihr Studium seltener abbrechen.* ÄrzteZeitung. Retrieved from https://www.aerztezeitung.de/Wirtschaft/Warum-Medizinstudenten-ihr-Studium-seltener-abbrechen-310940.html (last seen 12.04.2021).

Deutsche Telekom Stiftung. (2011). *Frühstudium*. Telekom Stiftung.

Deutsche Telekom Stiftung, Solzbacher, C. (2008). *Frühstudium—Schüler an die Universität*. Osnabrück.

Drews, P., & Martens, A. (2019). Juniorstudium—Study digital while going to school. In M. A. Herzog, Z. Kubincova, P. Han, M. Temperini (Eds.), *Advances in web-based learning ICWL 2019* (pp. 303–311).

Drews, P. & Martens, A. (2020). Digital support for the pre-student phase is successful. In *CELDA 2020, Online-Konferenz. IADIS* (pp. 133–140).

Fatzer, G. (1987). *Ganzheitliches Lernen. Humanistische Pädagogik und Pädagogik und Organisationsentwicklung.* Junfermann-Verlag.

Gröber, S., Müller, T., & Kuhn, J. (2018). Früheinstieg ins Physikstudium (FiPS)-Entwicklung der Konzeption eines Frühstudiums als Fernstudium. *PhyDid A-Physik und Didaktik in Schule und Hochschule, 1*(17), 1–12.

Heublein, U., Ebert, J., Hutzsch, C., Isleib, S., König, R., Richter, J., & Woisch, A. (2017). Zwischen Studienerwartungen und Studienwirklichkeit. *Forum Hochschule, 1*(1), 318.

Perleth, C., & Neumann, M. (2011). Vergleich von Präsenzlehrveranstaltungen mit inhaltsgleichen E-Learning-Angeboten im Lehramtsstudium. In *GI-Jahrestagung* (p. 454).

Solzbacher, C. (2011). Frühstudium in Deutschland: Ergebnisse einer bundesweiten Untersuchung. In *Beiträge zur Hochschulforschung-Themenheft "Frühstudium für begabte Jugendliche* (pp. 8–25).

Springermedizin. (2020). *Deutlich mehr Bewerber als freie Medizinstudienplätze.* Springermedizin. Retrieved April 12, 2021, from https://www.springermedizin.de/wissen-macht-arzt/deutlich-mehr-bewerber-als-freie-medizinstudienplaetze/18365166

Stanley, J. C. (1996). In the beginning: The study of mathematically precocious youth. In *Intellectual Talent: Psychometric and Social Issues* (pp. 225–235).

Triebner, S. (2010). *Randomisierte Studie zum Vergleich von zwei E-Learningformaten mit einer Präsenzveranstaltung*. Inaugural-dissertation, Friedrich-Alexander-Universität Erlangen-Nürnberg, Medizinischen Fakultät.

Chapter 8
Investigating How Word Clutter and Colour Impact Upon Learning

Olivia Foulds

8.1 Introduction

8.1.1 Defining Clutter

During human development, vision is fundamental to facilitate learning, knowledge and cognition (Chokron & Dutton, 2016). However when too much visual stimuli occur, human cognitive capacity is overloaded. Specifically, the detrimental effect of excess visual stimuli has been defined by the concept of clutter—otherwise known as crowding—which refers to the negative impact of nearby contours that interfere with and reduce visual discrimination when trying to focus on a target (Levi, 2008). Many different approaches have sought to precisely measure clutter in the hope that visual scenes can be assessed for the optimum amount of clutter present that doesn't affect an individual's perception (for a review, see Rosenholtz, Li, & Nakano, 2007). However, quantifying clutter is still difficult, as a recent review identified that many different stimuli can consist of clutter, including abstract symbols, letters and objects and even natural scenes (Patro & Huckauf, 2019).

Regardless of the preferred model for measuring clutter, theories have attempted to explain why clutter causes perceptual problems. Studies have revealed that when people view clutter compared to uncluttered scenes, the excess of visual stimuli exceeds the limits of attentional resources and short-term memory, which results in a bottleneck that impairs object perception (Levi, 2008). These results correlate well with research that identified different brain regions implicated in memory are activated when clutter is present (Hegdé, Thompson, Brady, & Kersten, 2012). Furthermore, eye-tracking measurements have revealed that when observing clutter,

O. Foulds (✉)
University of Strathclyde, Glasgow, Scotland, UK
e-mail: olivia.foulds@strath.ac.uk

D. Ifenthaler et al. (eds.), *Orchestration of Learning Environments in the Digital World*, Cognition and Exploratory Learning in the Digital Age,
https://doi.org/10.1007/978-3-030-90944-4_8

eyes fixate more peripherally resulting in slower and more methodical searching, confirming that attention is automatically diverted to clutter (Moacdieh & Sarter, 2015). Overall, the effects of clutter result in individuals being cognitively overwhelmed, which causes many negative outcomes, such as taking longer performing search tasks, yet obtaining less accurate results, and consequently learning less (Adamo, Cain, & Mitroff, 2015; Foulds, Azzopardi, & Halvey, 2021; Moacdieh & Sarter, 2015; Yeh, He, & Cavanagh, 2012).

8.1.2 Cluttered Learning Environments

In the last few years, educational research has begun to highlight the importance of visual crowding during learning. Although not specifically mentioning clutter, Fisher, Godwin, and Seltman (2014) claimed to be the first study that directly examined whether the visual environment of a classroom affected children's attention allocation. Children, with a mean age of 5 years, completed lessons where the classroom interchanged between sparsely decorated walls (no clutter) and heavily decorated visual displays (high clutter). Eye gaze revealed that in the decorated classroom, children were more distracted, and their understanding of their lesson was significantly lower (learning score mean = 42%) compared to learning in the clutter-free classroom (learning score mean = 55%).

Although Fisher et al.'s (2014) study represents a small sample of ages from just one American private nursery, similar results of the negative impact of clutter during learning have been found elsewhere. For example, in Hanley et al. (2017), a broader age range of English children (5–13 years old) participated in an experiment where they watched video lessons with a teacher that presented lesson content amongst a blank background (no clutter), whereas other lessons featured a background that contained a high amount of visual display (high clutter). Using eye-tracking measures and questioning the children about the content of the lessons, the findings replicated Fisher et al. (2014) and identified that the presence of visual clutter was associated with children attending to the background more, while retaining less information (Hanley et al., 2017). Similarly, in research using a more qualitative methodology, interviews with teachers and paraeducators highlighted that decluttering classrooms made a positive effect on students' learning experiences and behaviour (McDowell & Budd, 2018). With a range of different research methodologies finding similar results, the importance of clutter during education has been amplified.

8.1.3 Visual Crowding During Reading

When exploring the effect of clutter during education, it is worth investigating beyond the infrastructure of a classroom to also consider the materials that are presented to the learner. For example, when reading words, the inter-letter spacing used has been linked to clutter effects. More specifically, when spacing between letters was increased (compare education vs. e d u c a t i o n), participants had faster response times distinguishing between words or nonsense words (Perea, Moret-Tatay, & Gómez, 2011). This result was assumed to be a cause of increased spacing reducing the amount of clutter that surrounded each letter. However, other studies have found contrasting results. van den Boer and Hakvoort (2015) presented children who were beginning and advanced readers with word lists composed of a variety of inter-letter spacings (−0.5, default, 0.1, 1, 1.5, 2). Children were tested on their reading aloud fluency, and results found that increased inter-letter spacing did not benefit reading fluency for either beginning or advanced readers. However, inter-letter spacing that was smaller than default did significantly impair reading performance for both age groups (van den Boer & Hakvoort, 2015). This would imply that if letters are too close together, a clutter effect does arise, consequently impairing reading and thus minimising learning.

However, despite many studies exploring the effects of clutter on inter-letter spacing, to the best of my knowledge, no studies have yet directly investigated whether the effects of clutter extend onto more everyday reading scenarios, such as reading words that are surrounded by other words. Yet in the current era of technology, apps have been designed to provide alternative ways to read, such as Spritz (2019), VelocityReading (2020) and Look (CVI Scotland, 2021). These apps display text in a rapid serial visual presentation (RSVP): words do not appear in a block of sentences or paragraphs as they would in a normal book, but instead are quickly, sequentially presented one by one. It is believed that using RSVP methods results in faster reading times because an individual's eyes do not have to move (Spritz, 2019), which is why RSVP was traditionally designed as a mechanism to study neural activation without eye movement interrupting event-related potential responses (Kornrumpf, Niefind, Sommer, & Dimigen, 2016). Consequently, except one recent paper which suggests that RSVP may reduce clutter (Koornneef, Kraal, & Danel, 2019), little empirical research evidence has actually explained whether RSVP does reduce the negative effects of clutter. Additionally, when considering the focus of education and learning, if RSVP increases reading speed, it is unknown as to how well the words that are read at such faster speeds are retained. Therefore, if the phenomenon of clutter increases errors made in search tasks, limits short-term memory and increases reading time when letters are too crowded together, perhaps the results of reading through RSVP are also indicative of the phenomenon of clutter: words in a paragraph may be muddling the identification of the specific word that the reader wishes to focus on.

8.1.4 Colour

Generally, the negative perceptual effects of clutter have been shown to increase depending on the similarity of colour, contrast and visual complexity between the target and clutter that is present (Cheung & Cheung, 2017). Therefore, when participants were asked to identify a target letter surrounded by other irrelevant letters, the effects of clutter were found to reduce when the target and clutter differed in colour (Põder, 2007). Specifically, Põder's (2007) three experiments demonstrated that when distractor letters were black and the target letter was either red or highlighted in red or yellow, identification times of the target were significantly faster for all participants. Põder (2007) suggested that these results further accentuated clutter theories which state that excess visual stimuli exceeds the limits of attentional resources—by highlighting a target in a particular colour amongst cluttered distractors of a different colour, exogenously controlled attention was automatically attracted to the salient colour, thus reducing the cluttered illusion from the distractors. Although these experiments have not been replicated and the sample tested was very small, the initial findings from Põder (2007) would appear to suggest that colour does have the ability to affect reading of letters. This is in line with other research irrespective of clutter that has found faster reading speeds when every alternate word in a sentence was displayed in a different colour (Zhou, Ye, & Yan, 2020) or when coloured overlays have been used (Wilkins, Jeanes, Pumfrey, & Laskier, 1996). Similarly, coloured overlays have been used as a clinical intervention for individuals with dyslexia and autism to improve their attention, which demonstrates the power that colour can have on reading (Wilkins, 2003, cited in Dzulkifli & Mustafar, 2013; Ludlow & Wilkins, 2009).

Related to the theory that clutter doesn't just exceed attentional resources but is also implicated in memory (Levi, 2008), colour has also been identified as impacting upon generalised memory recall. For example, research spanning back over decades identified that participants remembered colour photographs significantly better than black-and-white images (Boyatziz & Varghese, 1999, cited in Olurinola & Tayo, 2015; Carter & Hartley, 2021); learners who employ colour coding had higher retention rates than conventional learning (Ozcelik et al., 2009, cited in Hutanu & Bertea, 2019); and a review on the effect of colour on memory concluded that colour can increase memory on certain tasks depending on the choice of colours used (Dzulkifli & Mustafar, 2013). For instance, when recalling letters in an experiment, slides with a white background elicited higher immediate retention rates than slides against a blue and green background (McConnohie, 1999, cited in Dzulkifli & Mustafar, 2013). Furthermore, Greene et al. (1983, cited in Dzulkifli & Mustafar, 2013) explained that colours such as red and orange elicited greater attentional capture compared to colours like brown and grey and that this was contributing to enhanced memory performance. Yet despite converging evidence indicating that colour impacts upon memory, little is known as to whether the colour of words affects memory during reading and learning in response to visual clutter.

8.1.5 Summary and Hypotheses

Overall, clutter increases response times and the number of errors made in search tasks (Moacdieh & Sarter, 2015); clutter in educational environments distracts learners, reduces understanding (Fisher et al., 2014) and results in children retaining less information (Hanley et al., 2017); and clutter of individual letters impairs reading (van den Boer & Hakvoort, 2015). However, little is known for whether the detrimental effects of clutter extend onto an individual's perception when reading and learning words that are surrounded by other words. In accordance with the measure of clutter proposed by Yu, Samaras, and Zelinsky (2014), the present study quantified clutter by the amount of similar image features that merge into a perceptual scene. Thus, when reading words, 'clutter' was defined by multiple irrelevant words and non-words visible on the same screen. It was firstly hypothesised that words learned amongst either form of clutter would be associated with worse and slower accuracy in a recognition task, compared to words learned in isolation. Furthermore, as research has identified the colour red as eliciting greater attention and faster reading times (Dzulkifli & Mustafar, 2013; Põder, 2007), words that are learned in red font should hypothetically elicit greater and faster accuracy than words learned in black font. Finally, since Põder (2007) established that colour interacts with clutter, such that red letters amongst black distractors reduced the negative perceptual effects of clutter, it was hypothesised that target words learned in red font amongst clutter would result in similar post-task recognition accuracy and reaction times to words that were learned in isolation.

In summary, the hypotheses defined for this study are:

- H1: Target words learned amongst distractor words will be remembered less accurately and slower in a recognition task, compared to target words learned in isolation.
- H2: Target words learned in red font will be remembered more accurately and faster in a recognition task, compared to target words learned in black font.
- H3: Target words learned in red font amongst clutter will produce similar accuracy and reaction times to target words that were learned in isolation.

8.2 Methodology

8.2.1 Design

An experiment was conducted using a 3 (clutter) × 2 (colour) repeated-measures factorial design. Clutter was defined by what surrounded the target word that participants were instructed to remember. This resulted in three levels, with target words being presented in (1) isolation (no clutter), (2) with eight irrelevant words around each target (clutter words) and (3) eight irrelevant non-words around each target (clutter non-words). Non-words were used to ensure that any results found were in fact due to clutter and not just related to the linguistic structure that distractor words would automatically be read by the reader. Colour was operationalised so that target words were either presented in standardised black or red font. For the recognition task, subjects were then presented with a word and asked to identify whether it had been a target they had already seen. Therefore, average reaction

Table 8.1 The experimental conditions

Target word colour	No clutter	Clutter words	Clutter non-words
Black	Condition 1	Condition 2	Condition 3
Red	Condition 4	Condition 5	Condition 6

times and correct responses for the recognition task in each condition were measured as the dependent variables. Overall, every participant partook in six conditions (Table 8.1):

To ensure a fair comparison between the conditions, the number of letters, syllables and specific properties of words were equally balanced across all conditions. To reduce any order or practice effects from occurring, the order of condition presentation was randomised so that every participant viewed the conditions in a unique sequence. Finally, the entire experiment was designed to be compatible with the British Psychological Society (2018) ethical guidelines.

8.2.2 *Participants*

Forty-two adults ranging from 19 to 78 years old participated (29 female; 13 male; mean age = 42.6 years). Participants were recruited through a university experimental participation website. All participants consented online to being able to read English, had normal or corrected-to-normal vision by self-report, were naïve to the purpose of the experiment and received no payment for participation.

8.2.3 *Materials*

The experiment was programmed and run using OpenSesame (Mathôt, Schreij, & Theeuwes, 2012). Stimuli consisted of target words to remember (10 per condition), distractor words for the clutter-word conditions (8 per target), distractor non-words for the clutter-non-word conditions (8 per target) and additional ‘foil’ words for each recognition task (20 per condition). All stimuli were presented in 32-pt font with default inter-letter spacing against a white background. All words were selected through generating a list of items on the English Lexicon Project database (Balota et al., 2007).

Target words, which were always centrally positioned in either standardised black or red font, were chosen to ensure each condition had an equivalent word for word length, number of syllables, beginning letter, Freq_HAL > 1000 and a I_Mean_RT range of 500–900. Plural and distressing words were excluded, and words ranged from four to eight letters long, comprising either one or two syllables.

Similar to the target words, the distractor words for the clutter-word conditions ranged from four to seven letters long, with one or two syllables, and each clutter condition had an equivalent word such as ‘Turkey’ and ‘Turtle’.

The distractor non-words for the clutter-non-word conditions were created by rearranging letters from each clutter word, ensuring that each non-word had an equivalent normal word with regard to beginning letter, word length and syllabic structure and looked linguistically pronounceable. For example, ‘tennis’ was changed to ‘tesnin’.

Both distractor words and distractor non-words always appeared in black font that symmetrically surrounded the target, depending on the condition (see Fig. 8.1).

The additional foil words for the recognition task, always presented centrally in black font, were as alike to the target words as possible by sharing the same number of letters, syllables, beginning letter and frequency range.

The order of word presentation in each condition was randomised. Furthermore, a Python code inline script was executed to ensure randomisation of condition order. Additionally, words were never repeated between conditions. To measure responses, a standard QWERTY keyboard was used. Data was then individually saved into Microsoft Excel outputs and transferred into SPSS software for statistical analysis.

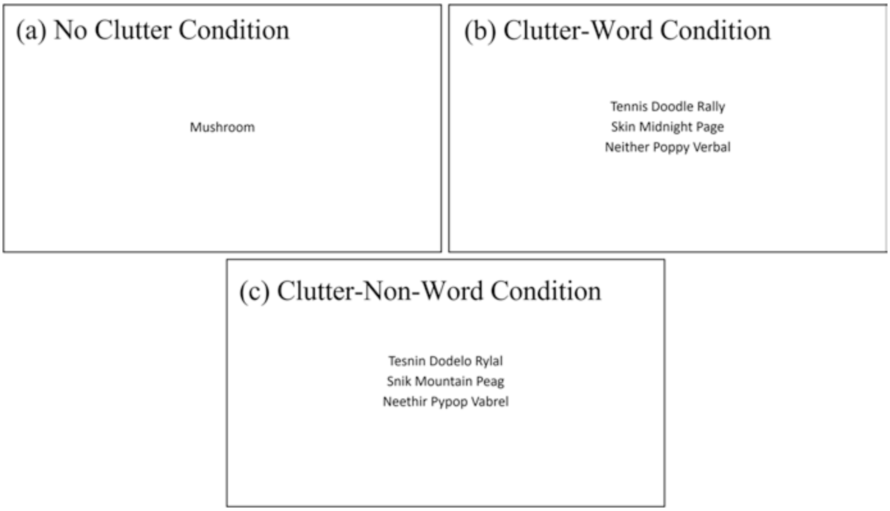


Fig. 8.1 Examples of how stimuli were presented in different conditions when the target word was presented in black font. In the red font conditions, the target word (such as Mushroom, Midnight and Mountain) would have been in red font

8.2.4 Procedure

The experiment began with an information page where participants were informed they were partaking in a memory experiment. Then participants provided their gender and age and gave online informed consent. Instructions explicitly informed participants to ensure they were sitting at a comfortable viewing distance from the computer screen, in a quiet room, with no distractions. Additionally, the instructions told participants to remember words that were presented in the centre of the screen, where a fixation point would direct them to look. Participants were also informed that the word they were to remember would sometimes be presented in isolation, and at other times would have other words visible on the same screen, and that they were only to attend to the central word, where the fixation point had been.

A practice trial familiarised participants with the task, and these scores were not included in the analyses. The practice trial consisted of a central target word presented against eight other distractor words, followed by another target word, both visible for 6000 ms. A screen then appeared that informed participants they would see another list of words (where each word would be presented on the screen individually), and that they had to respond whether they had been one of the initial target words by pressing 'z' for yes and 'm' for no or to press nothing if they did not know and wait for the screen to change.

After the practice trial, there were six experimental blocks/conditions to complete, which each participant completed in a random order. For each block, ten target words were sequentially positioned in the centre of the screen for 6000 ms each. Between each target word presentation, a 500 ms fixation point appeared. In the clutter conditions, eight words or non-words would symmetrically surround the target word. For the recognition task, the same target words were randomly interspersed with an additional 20 words, and these would appear visible until participants pressed 'z' or 'm' or until 5000 ms had elapsed. There was no feedback about the correctness of responses.

Between each block, participants were informed that they no longer needed to remember the previous words and that they could initiate the next block whenever they were ready by pressing any key. The same instructions were repeated between blocks to remind participants what to do.

Overall, the experiment lasted about 20 min. Afterwards, a virtual debriefing statement was displayed that thanked individuals for participating, explained the purpose of the experiment and provided the researcher's contact details for any future concerns they may have had.

8.3 Results

Multiple repeated-measures ANOVAs were conducted to compare the accuracy and reaction times to identifying target words in a recognition task, depending on whether they had been learned in isolation (no clutter, NC), amongst clutter words (CW), amongst clutter non-words (CNW) or in black or red font.

8.3.1 Accuracy

The results of a 3 × 2 repeated-measures ANOVA with a Greenhouse-Geisser correction revealed that there was a significant main effect of clutter on participant accuracy at identifying target words ($F(1.67, 68.28) = 9.10$, $p = 0.001$, $\eta_p^2 = 0.182$). Bonferroni Post Hoc tests confirmed that mean accuracy scores were significantly higher when identifying target words that had been learned in no clutter (NC = 8.92 targets correctly identified) compared to target words that had been learned amongst clutter words (CW = 7.93) and clutter non-words (CNW = 8.13). There was no significant difference between accuracy on CW and CNW.

While descriptive statistics revealed that participant mean accuracy scores were slightly higher when verifying target words that had been learned in red font (8.49) compared to black font (8.16), the ANOVA with sphericity assumed revealed that this difference was not significant ($F(1,41) = 3.93$, $p = 0.054$, $\eta_p^2 = 0.087$).

Furthermore, there was no significant interaction between clutter and colour ($F(2,82) = 1.07$, $p = 0.346$, $\eta_p^2 = 0.026$). However, for each clutter condition, a trend emerged that participants did have slightly higher accuracy when the target word had been initially presented in red as opposed to black font (see Table 8.2 and Fig. 8.2).

Table 8.2 Descriptive statistics (mean ± standard deviation) for accuracy as the number of correctly identified target words per condition

	Accuracy	
	Target word colour	
Clutter type	Black	Red
No clutter (NC)	8.76 ± 1.66	9.07 ± 1.13
Clutter words (CW)	7.62 ± 2.00	8.24 ± 2.26
Clutter non-words (CNW)	8.10 ± 1.82	8.17 ± 1.85

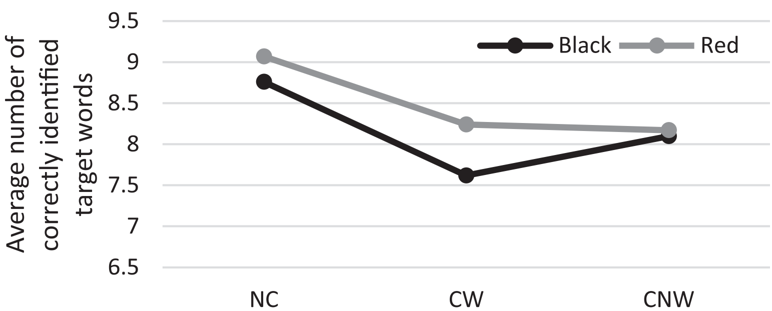


Fig. 8.2 Accuracy for the number of correctly identified target words for each condition, where black and red refer to the colour of the target word

8.3.2 Screened Reaction Time

Reaction times (RTs) to incorrect target word responses were treated as missing data and were screened and replaced by predictive mean matching for the average RT for each participant's correct responses. While descriptive statistics revealed that participant mean reaction times for target words that had been learned in no clutter were slightly faster (NC = 895.78 ms) than words that had been learned in clutter (CW = 903.74 ms; CNW = 905.74 ms), a repeated-measures ANOVA with Greenhouse-Geisser correction revealed that this difference was not significant ($F(1.97, 803.28) = 0.14$, $p = 0.865$, $\eta_p^2 = 0.000$).

Additionally, the results of the ANOVA with sphericity assumed revealed that there was no significant main effect of target word colour on reaction times ($F(1,408) = 0.01$, $p = 0.922$, $\eta_p^2 = 0.000$). Participant mean reaction times for identifying black target words (900.99 ms) were very similar to that of red target words (902.52 ms).

Furthermore, there was no significant interaction for reaction times between clutter and colour ($F(2, 816) = 2.74$, $p = 0.065$, $\eta_p^2 = 0.007$). However, the responses for red and black target words did differ depending on the clutter condition: when no clutter was present, black target words elicited faster reaction times than red target words; when clutter words were used, black target words elicited slower reaction times than red target words; and when clutter non-words were used, similar to the no-clutter condition, black target words elicited faster reaction times than red target words (see Fig. 8.3 for an illustration of the means).

8.3.3 Raw Reaction Time

To provide a more in-depth examination of data, when reaction times (RTs) to incorrect target word responses were included in analyses (following the protocol of Yeh et al., 2012), a repeated-measures ANOVA with Greenhouse-Geisser correction

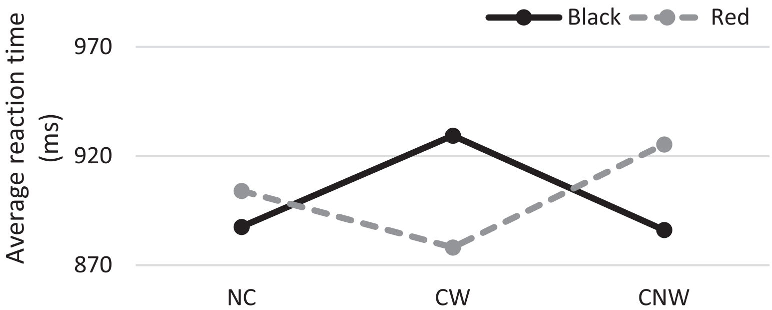


Fig. 8.3 Average reaction times (ms) for target word recognition responses for each condition, analysed with predictive mean matching imputation for any incorrect responses. Black and red refer to the target word colour

did identify a significant main effect of clutter on reaction time (F(1.94,2384.28) = 11.35, p = 0.000, η_p^2 = 0.009). Post hoc tests (Bonferroni) revealed that reaction times to target words presented in no clutter (NC = 887.25 ms) were significantly faster than both forms of clutter (CW = 960.60 ms; CNW = 940.82 ms).

Although reaction times for the red words were slightly faster (927.82 ms) than the black words (931.29 ms), this result did not reach significance (F(1,1229) = 0.07, p = 0.789, η_p^2 = 0.000).

However, in comparison to the interaction between clutter and colour for screened RTs, raw RTs produced the reverse pattern: when there was no clutter or clutter non-words, participants responded slightly faster to red words than black words; yet when clutter words were present, participants responded slightly faster to black words than red words (see Fig. 8.4 and Table 8.3 for an illustration of the means). Nonetheless, this interaction between clutter and colour did not reach significance (F(1.99, 2440.43) = 1.82, p = 0.162, η_p^2 = 0.001).

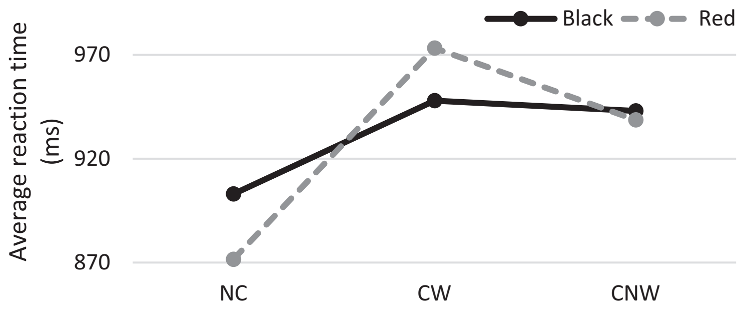


Fig. 8.4 Average reaction time (ms) for target word responses for each condition, including incorrect responses. Black and red refer to the target word colour

Table 8.3 A comparison between raw and screened data for reaction times in ms (mean ± standard deviation) in each condition

Condition	Screened data	Raw data
No clutter: Black	887.52 ± 387.03	902.96 ± 517.82
No clutter: Red	904.05 ± 468.80	871.54 ± 418.92
Clutter words: Black	929.37 ± 433.94	947.94 ± 624.94
Clutter words: Red	878.11 ± 372.12	973.26 ± 745.75
Clutter non-words: Black	886.09 ± 385.82	942.98 ± 577.62
Clutter non-words: Red	925.39 ± 465.93	938.67 ± 627.65

8.4 Discussion

As predicted, words that were learned in clutter elicited significantly impaired accuracy to recalling them in a recognition task, compared to words that were learned in isolation. Similarly, reaction times to identifying target words that had been learned in clutter were generally slower compared to no clutter—although this pattern only reached significance when incorrect responses were included in analyses. Additionally, despite the hypothesis that target words learned in red font would result in greater accuracy and faster reaction times compared to black target words, no significant differences were found. However, a trend did appear that red target words elicited higher recognition accuracy and were recalled slightly faster than black words. Furthermore, while it was predicted that red target words presented amongst black clutter might reduce the negative effects of clutter and lead to similar accuracy and reaction times to words that had been learned in isolation, there was no significant interaction found between clutter and colour. However, an interesting pattern emerged that when participants' incorrect answers for recognition accuracy were included in further analyses, red target words elicited the fastest reaction times in the no-clutter and clutter-non-word condition. Yet, when incorrect responses were excluded from analyses, black target words elicited the fastest reaction times in the no-clutter and clutter-non-word condition.

The finding that clutter around a target word, either in the form of irrelevant words or non-words, increased the number of errors made and reaction times in response to a recognition task, would appear consistent with a multitude of previous research that identified clutter as degrading accuracy and response times in search tasks (Adamo et al., 2015; Moacdieh & Sarter, 2015; Yeh et al., 2012). Additionally, the present findings assimilate to faster reading times when the clutter surrounding an individual letter in a word was reduced (Perea et al., 2011; van den Boer & Hakvoort, 2015). As the present study found a significant impairment of target word recognition when the word had been learned in clutter as opposed to isolation, the most reasonable conclusion for these results would appear to relate to other study's explanations of clutter—negative perceptual effects of clutter do extend onto more everyday reading scenarios, such as words surrounding other words: clutter automatically attracts vision away from the target, and this excess of visual stimuli exceeds the limits of attentional resources and short-term memory (Levi, 2008). For example, when participants are attempting to memorise a target word, if other words or non-words are surrounding the target, these may be mistakenly processed as the target word and cause a degradation of memory, because only a certain number of items can be processed at one time (Chastain, 1991).

Furthermore, an interesting observation with the present study is that participants were explicitly pre-warned that they did not need to look at the words beyond the centre of the screen. Therefore, it cannot be concluded that attention was diverted to clutter because participants were not expecting it. Thus, it is unknown whether the effects of clutter are so strong that they overrule overt attention and distract a user's vision regardless of whether they want to process it or not, or whether the effects may be a result of more subconscious and covert behaviour that participants are unaware of, perhaps through their peripheral vision subtly processing the excess stimuli. Consequently more research is needed, perhaps using eye tracking, to explain how exactly visual attention responds to clutter during reading.

In keeping with theories related to attention, colour has previously been thought to capture attention and thus enhance memory performance (Dzulkifli & Mustafar, 2013). Although the present study did not find a significant effect of learning words in the colour red, compared to black, on accuracy of identifying target words in a recognition task, this finding only marginally missed out on reaching significance ($p = 0.054$). Therefore, it could be tentatively assumed that a larger sample size may have resulted in a significant result and that words learned in red font do in fact elicit better recognition than black. This explanation would reiterate the findings by Greene et al. (1983, cited in Dzulkifli & Mustafar, 2013), who stated that red elicited greater attentional capture compared to colours like brown and grey and that this impacted upon memory performance. However, other research has claimed that long-wavelength colours, like red, actually illicit worse performance in certain tasks (Hall & Hanna, 2004; Shieh & Lin, 2000). Furthermore, unlike research that identified colour as increasing reading speeds (Wilkins et al., 1996), the present study found minimal differences in reaction times for either red or black target words. Thus, further research is needed to clarify the contrasting results of whether the colour red enhances or hinders individual's performance in tasks.

It is unsurprising that as the present study found no significant benefit of presenting the target word in red, compared to black, red target words also failed to reduce the negative effects of surrounding clutter. These results are in contrast to Põder's (2007) experiments that found when target and surrounding clutter differed in colour, the process of mentally eliminating distractors was easier (Levi, 2008). However, these findings should not automatically reject the theory that colour can reduce the negative effects of clutter because only two colours were used, red and black; participants were not screened for colour blindness; and due to variations in computer browser that the experiment was viewed on, the brightness could not be controlled for across participants. Additional examination using different methods would be useful to clarify these methodological problems of how the colour red contrasted to the colour black.

Nonetheless, the current research did identify an interesting interaction in reaction times: when incorrect responses were excluded from analyses, reaction times were faster for identifying black target words in the no-clutter and clutter-non-word condition and faster for red target words in the clutter-word condition. Yet, when incorrect responses were included in analyses, the opposite pattern emerged. These results would imply that when participants don't know the correct answer, target words learned in red font generally elicited a faster response in the recognition task compared to black words. Contrastingly, when participants did know the answer, words learned in black font were responded to the fastest. The exact mechanism behind these findings is unknown, and as this interaction did not reach statistical significance, this pattern could be purely the result of chance. However, it would seem worthwhile further investigating whether the interaction between colour and clutter is in fact more complex than previous theories have identified.

The results of the present experiment are not just theoretical, but could also have vital importance in many aspects of everyday life, because the finding that clutter does impair individual's performance across a multitude of tasks has been replicated. More specifically, the present study provided the first evidence (to the best of my awareness) that the negative perceptual effects of clutter do extend onto reading words surrounded by other words. As this is how many textbooks and websites present information, the present experimental findings could evolve current learning strategies and alter popular opinion on how best to read and retain words. With vision being a huge facilitator of learning (Chokron & Dutton, 2016), it may be a more positive and beneficial learning experience if clutter is considered for future design of learning materials. Therefore apps that operate rapid serial visual presentation, or handheld devices like typoscopes which ensure only a few words are visible at any one time, could potentially offer a more optimal reading environment in comparison to a standard book. This alternative way to read should then hypothetically promote greater memory of what has been read and in a faster time period. Presenting words in a minimalistic structure may also enable more inclusive education to occur, as other research has identified that children with cerebral visual impairment or dyslexia particularly struggle with clutter (Chokron & Dutton, 2016; Morris, Fourney, Ali, & Vonessen, 2018).

However, it is worth noting some limitations of the present study. Firstly, in contrast to clutter studies conducted in a controlled laboratory, anyone with OpenSesame could participate in the experiment. Therefore, although all words were presented in a consistent size, due to variation amongst computers and distance from the screen, some participants would have seen the words in different sizes, resolutions and colour contrasts. Other studies on clutter have removed this confound and ensured a fixed viewing distance by using a chin rest (Adamo et al., 2015; Yeh et al., 2012). However, this may not be necessary when other research has concluded that clutter perception is invariant to image size, but dependent on the content (Zelinsky & Yu, 2015). Furthermore, as the experiment was conducted virtually, it was unknown whether participants varied in terms of factors that may affect their memory such as their mood, the time of day the experiment was completed, distractions that were present and alcohol or drug consumption. Nonetheless, these limitations could in fact be seen as a positive—no data was excluded from analyses, yet results were still significant. Therefore, the overall finding that clutter does significantly impair word recognition may extend onto a population irrelevant of individual situation or distractions present.

In order to fully establish whether the effects of clutter similarly extend onto other populations, further research would firstly be needed. Beyond individual differences, it would appear worthwhile to examine whether the degradation of word recognition amongst other distractor words exists amongst children—after all, it is children who are expected to memorise what they learn through textbooks and websites that contain a plethora of clutter. It may be possible that children are unaffected by clutter and that it is a concept that develops and worsens with age. Alternatively, it is unknown whether the negative effects of clutter can be alleviated through practice, and therefore reading words surrounded by other words in a normal paragraph do not result in clutter effects if readers are familiar with reading in that way. Correspondingly, as the present experiment only tested recognition of individual words, it would be interesting to expand the ecological validity and see whether clutter affects word recognition at a sentence level. Furthermore, future research should develop the design of the current experiment to explore whether learning words amongst clutter impairs memory recognition beyond the immediate short-term memory and not just instantly after words have been presented. Consequently, the present research is only the beginning of a topic that could potentially improve learning in many domains. All of these unanswered questions therefore raise many interesting avenues for future research that could have hugely important real-life consequences for how best to learn when clutter is present. This will provide evidence that will steer practical guidelines for a variety of domains, such as creating clutter-free websites, learning resources and information retrieval platforms. If more understanding of clutter is had, then hopefully strategies can be implemented to reduce the negative effects it causes.

8.5 Conclusion

To summarise, the current experiment identified that clutter, either in the form of distractor words or non-words, significantly impaired accuracy and increased reaction times for recognising target words. This finding would appear to extend other theories that when excess visual stimuli are present, the limits of short-term memory and attention are exceeded. Although no significant findings were found in relation to the colour of a target word or the interaction between clutter and colour, some interesting patterns and trends emerged that require more investigation. Finally, as this experiment provides the first evidence that clutter effects do extend onto words surrounding other words, these findings motivate further research for creating an optimal reading format for learning. For example, if clutter can be reduced during reading, through ways such as apps that operate rapid serial visual presentation, then reading speed should be faster while improving memory for what has been seen.

Acknowledgements This chapter was revised from work undertaken as part of the author's undergraduate degree with The Open University. The author would like to thank their tutor at the time, M. Louise Kelly, for their original feedback. Additionally, appreciation goes to the anonymous reviewers for providing helpful suggestions.

References

Adamo, S. H., Cain, M. S., & Mitroff, S. R. (2015). Targets need their own personal space: Effects of clutter on multiple-target search accuracy. *Perception, 44*(10), 1203–1214. https://doi.org/10.1177/0301006615594921

Balota, D. A., Yap, M. J., Hutchison, K. A., Cortese, M. J., Kessler, B., Loftis, B., Neely, J. H., Nelson, D. L., Simpson, G. B., & Treiman, R. (2007). The English lexicon project. *Behavior Research Methods, 39*(3), 445–459.

British Psychological Society. (2018). *Code of ethics and conduct* [Online]. The British Psychological Society Ethics Committee. Retrieved from https://www.bps.org.uk/sites/www.bps.org.uk/files/Policy/Policy%20-%20Files/BPS%20Code%20of%20Ethics%20and%20Conduct%20%28Updated%20July%202018%29.pdf

Carter, C. K., & Hartley, C. (2021). Are children with autism more likely to retain object names when learning from colour photographs or black-and-white cartoons? *Journal of Autism and Developmental Disorders, 51*(9), 3050–3062. https://doi.org/10.1007/s10803-020-04771-2

Chastain, G. (1991). Effects of abruptly appearing clutter on a peripherally precued covert attention shift. *The Journal of General Psychology, 118*(1), 31–44. https://doi.org/10.1080/00221309.1991.9711131

Cheung, L. Y. T., & Cheung, S. H. (2017). Chinese-character crowding—I. Effects of structural similarity. *Journal of Vision, 17*(11), 1–13. https://doi.org/10.1167/17.11.14

Chokron, S., & Dutton, G. N. (2016). Impact of cerebral visual impairments on motor skills: Implications for developmental coordination disorders. *Frontiers in Psychology, 7*(1471), 1–15. https://doi.org/10.3389/fpsyg.2016.01471

CVI Scotland. (2021). *Look* [Computer application]. Retrieved from https://cviscotland.org/utils/look/

Dzulkifli, M. A., & Mustafar, M. F. (2013). The influence of colour on memory performance: A review. *The Malaysian Journal of Medical Sciences: MJMS, 20*(2), 3–9.

Fisher, A. V., Godwin, K. E., & Seltman, H. (2014). Visual environment, attention allocation, and learning in young children: When too much of a good thing may be bad. *Psychological Science, 25*(7), 1362–1370. https://doi.org/10.1177/0956797614533801

Foulds, O., Azzopardi, L., & Halvey, M. (2021, March). Investigating the influence of ads on user search performance, behaviour, and experience during information seeking. In *Proceedings of the 2021 Conference on Human Information Interaction and Retrieval* (pp. 107–117). https://doi.org/10.1145/3406522.3446024.

Hall, R. H., & Hanna, P. (2004). The impact of web page text-background colour combinations on readability, retention, aesthetics and behavioural intention. *Behaviour & Information Technology, 23*(3), 183–195. https://doi.org/10.1080/01449290410001669932

Hanley, M., Khairat, M., Taylor, K., Wilson, R., Cole-Fletcher, R., & Riby, D. M. (2017). Classroom displays—Attraction or distraction? Evidence of impact on attention and learning from children with and without autism. *Developmental Psychology, 53*(7), 1265. https://doi.org/10.1037/dev0000271

Hegdé, J., Thompson, S. K., Brady, M., & Kersten, D. (2012). Object recognition in clutter: Cortical responses depend on the type of learning. *Frontiers in Human Neuroscience, 6*(170), 1–15. https://doi.org/10.3389/fnhum.2012.00170

Hutanu, A., & Bertea, P. E. (2019). A review of eye tracking in Elearning. In *Proceedings of the 15th International Scientific Conference eLearning and Software for Education*. https://doi.org/10.12753/2066-026X-19-038.

Koornneef, A., Kraal, A., & Danel, M. (2019). Beginning readers might benefit from digital texts presented in a sentence-by-sentence fashion. But why? *Computers in Human Behavior, 92*, 328–343. https://doi.org/10.1016/j.chb.2018.10.024

Kornrumpf, B., Niefind, F., Sommer, W., & Dimigen, O. (2016). Neural correlates of word recognition: A systematic comparison of natural reading and rapid serial visual presentation. *Journal of Cognitive Neuroscience, 28*(9), 1374–1391. https://doi.org/10.1162/jocn_a_00977

Levi, D. M. (2008). Crowding—An essential bottleneck for object recognition: A mini-review. *Vision Research, 48*(5), 635–654. https://doi.org/10.1016/j.visres.2007.12.009

Ludlow, A. K., & Wilkins, A. J. (2009). Case report: Color as a therapeutic intervention. *Journal of Autism and Developmental Disorders, 39*(5), 815–818. https://doi.org/10.1007/s10803-008-0672-5

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods, 44*(2), 314–324. https://doi.org/10.3758/s13428-011-0168-7

McDowell, N., & Budd, J. (2018). The perspectives of teachers and paraeducators on the relationship between classroom clutter and learning experiences for students with cerebral visual impairment. *Journal of Visual Impairment and Blindness, 112*(3), 248–260.

Moacdieh, N., & Sarter, N. (2015). Clutter in electronic medical records: Examining its performance and attentional costs using eye tracking. *Human Factors, 57*(4), 591–606. https://doi.org/10.1177/0018720814564594

Morris, M. R., Fourney, A., Ali, A., & Vonessen, L. (2018, April). Understanding the needs of searchers with dyslexia. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (pp. 1–12). https://doi.org/10.1145/3173574.3173609.

Olurinola, O., & Tayo, O. (2015). Colour in learning: Its effect on the retention rate of graduate students. *Journal of Education and Practice, 6*(14), 1–5.

Patro, K., & Huckauf, A. (2019). Asymmetries in flanker-target interference at different levels of number processing. *Acta Psychologica, 201*(102938), 1–11. https://doi.org/10.1016/j.actpsy.2019.102938

Perea, M., Moret-Tatay, C., & Gómez, P. (2011). The effects of interletter spacing in visual-word recognition. *Acta Psychologica, 137*(3), 345–351. https://doi.org/10.1016/j.actpsy.2011.04.003

Põder, E. (2007). Effect of colour pop-out on the recognition of letters in crowding conditions. *Psychological Research, 71*(6), 641–645. https://doi.org/10.1007/s00426-006-0053-7

Rosenholtz, R., Li, Y., & Nakano, L. (2007). Measuring visual clutter. *Journal of Vision, 7*(2), 17–17. https://doi.org/10.1167/7.2.17

Shieh, K. K., & Lin, C. C. (2000). Effects of screen type, ambient illumination, and color combination on VDT visual performance and subjective preference. *International Journal of Industrial Ergonomics, 26*(5), 527–536. https://doi.org/10.1016/S0169-8141(00)00025-1

Spritz. (2019). *Spritz* [Computer application]. Retrieved from http://spritzinc.com

van den Boer, M., & Hakvoort, B. E. (2015). Default spacing is the optimal spacing for word reading. *Quarterly Journal of Experimental Psychology, 68*(4), 697–709. https://doi.org/10.1080/17470218.2014.964272

VelocityReading. (2020). *Velocity reading* [Computer application]. Retrieved from https://velocityreading.com/learn

Wilkins, A. J., Jeanes, R. J., Pumfrey, P. D., & Laskier, M. (1996). Rate of Reading Test®: Its reliability, and its validity in the assessment of the effects of coloured overlays. *Ophthalmic and Physiological Optics, 16*(6), 491–497.

Yeh, S. L., He, S., & Cavanagh, P. (2012). Semantic priming from crowded words. *Psychological Science, 23*(6), 608–616. https://doi.org/10.1177/0956797611434746

Yu, C. P., Samaras, D., & Zelinsky, G. J. (2014). Modeling visual clutter perception using proto-object segmentation. *Journal of Vision, 14(7):4*, 1–16. https://doi.org/10.1167/14.7.4

Zelinsky, G. J., & Yu, C. P. (2015). Clutter perception is invariant to image size. *Vision Research, 116*, 142–151. https://doi.org/10.1016/j.visres.2015.04.017

Zhou, W., Ye, W., & Yan, M. (2020). Alternating-color words facilitate reading and eye movements among second-language learners of Chinese. *Applied PsychoLinguistics, 41*(3), 685–699. https://doi.org/10.1017/S0142716420000211

Chapter 9
Evaluating the Spatial Continuity Effects of Augmented Reality System on Learning Performance and Psychological Factors

Xuewang Geng and Masanori Yamada

9.1 Introduction

The development of high technology and the widespread use of smart mobile devices have not only changed our lives but also impacted education. Communication and graphics technologies have made various new learning methods possible (Nincarean, Alia, Halim, & Rahman, 2013). Augmented reality (AR) and virtual reality (VR) technologies provide learners with the opportunity to interact with the learning content and allow them to experience it directly (Ibáñez, Di Serio, Villarán, & Delgado-Kloos, 2016). In addition, the presence of smart mobile devices allows AR technologies to be experienced without the need to wear head-mounted displays since AR can now easily and effectively be experienced with smartphones. AR technology has been used in games, navigation, geological research, etc. and has great possibilities for application in education (Akçayır & Akçayır, 2017). The results of AR in education approaches have shown that AR increases learning motivation, improves satisfaction, makes abstract or complex concepts easier to understand, and facilitates learning (Cabero-Almenara, Fernández-Batanero, & Barroso-Osuna, 2019; Ibrahim et al., 2018; Santos et al., 2013; Walczak, Wojciechowski, & Cellary, 2006).

With the growing demand for language learning, there has been an increasing amount of literature on mobile-assisted language learning (MALL) (Cakmak, 2019). AR integrates haptic and visual learning content, allowing learners to combine more meaningful cues with learning content in the real world to facilitate

X. Geng (✉)
Graduate School of Human-Environment Studies, Kyushu University, Kyushu, Japan
e-mail: geng@mark-lab.net

M. Yamada
Faculty of Arts and Science, Kyushu University, Kyushu, Japan

D. Ifenthaler et al. (eds.), *Orchestration of Learning Environments in the Digital World*, Cognition and Exploratory Learning in the Digital Age,
https://doi.org/10.1007/978-3-030-90944-4_9

language acquisition (Santos et al., 2013). In MALL studies using AR technology approaches, AR has been shown to be more effective in learning than using text alone. For example, Santos et al. (2013) stated that AR learning through annotated objects was more effective than learning using paper textbooks. AR systems have the potential to improve learning performance and knowledge retention in language acquisition. Since the learning content can be verbal, pictorial, or hybrid, mobile AR language learning systems as a multimedia learning environment, multimedia learning design principles (Mayer, 2009) need to be considered. The principle of spatial continuity is the multimedia learning design principle. Spatial continuity design contributes to improving learning by reducing the extraneous cognitive load (Schroeder & Cenkci, 2018). This study evaluates AR learning systems designed using the spatial continuity principle by comparing learning performance, cognitive load, and technology acceptance.

This paper is organized as follows: Section 9.2 describes the literature review on AR systems in language learning and the spatial continuity principle. Section 9.3 introduces AR compound verb learning systems. Section 9.4 describes the study's research questions. Section 9.5 presents the methodology used in this study. Sections 9.6 and 9.7 present the results and discuss these findings. Finally, Section 9.8 presents the conclusions of the study.

9.2 Literature Review

9.2.1 Mobile AR Systems in Language Acquisition

MALL evolved from computer-assisted language learning (CALL) and mobile learning (m-learning) (Cakmak, 2019). Cakmak (2019) stated that MALL supports language learning using mobile technology and provides many "authentic," "relevant," and "contextual" language learning experiences. AR technology on mobile devices has a great potential to support language learning. AR technology has three features: a combination of real time and virtual, real-time interaction, and being registered in 3D (Azuma, 1997). These features bring the following advantages to AR in verbal learning: real-world annotation, contextual visualization, and visual-haptic visualization (Santos et al., 2013). AR enables learners to link learning content to reality scenarios for elaboration by integrating virtual objects into the real environment through real-world annotation and contextual visualization. Additionally, AR integrates perceived visual images and haptics, providing more modalities to present visual information. Data from several studies have suggested that AR language learning systems on mobile devices facilitate language acquisition (e.g., Boonbrahm, Kaewrat, & Boonbrahm, 2015; Hsu, 2017; Ibrahim et al., 2018; Santos et al., 2016). For example, Santos et al. (2016) have shown that handheld AR systems for foreign vocabulary learning improve knowledge retention.

AR systems can be divided into location-based and image-based AR systems (Wojciechowski & Cellary, 2013). Location-based AR systems are those in which the recognition function uses data from a mobile device location service (GPS or Wi-Fi positioning system). Learners can learn different learning content in different geographical locations. Image-based AR systems use image recognition technology to identify objects with virtual information attached to them. In this study, we used an image-based AR system to observe the interactions between learning materials and learners.

9.2.2 The Spatial Continuity Principle

As the learning contents of the mobile AR language learning system are in text, picture, animation, or hybrid form, the system's design needs to consider multimedia learning design principles (Mayer, 2009). According to the cognitive theory of multimedia learning, learners should have the cognitive ability to select information, organize relevant information, and integrate it into prior knowledge structures since human working memory is limited (Mayer, 2014). Mayer (2014) integrated cognitive load theory (Sweller, 2010) to clarify three types of cognitive load in multimedia learning environments: intrinsic cognitive load, extraneous cognitive load, and germane cognitive load. According to Sweller (2010), intrinsic cognitive load is determined by the underlying characteristics of the information in the learning content rather than by the instructional design, depending on the complexity of the learning material and the learner's prior knowledge. Extraneous cognitive load is related to the design of the material, and ineffective instructional design may lead to distractions that increase extraneous cognitive load. Meanwhile, germane cognitive load refers to the necessary tasks that aid learning and is related to the acquisition of knowledge.

The spatial continuity principle is an instructional design technique developed using the cognitive theory of multimedia learning on a theoretical basis (Schroeder & Cenkci, 2018). The use of the principle of spatial continuity also needs to be considered in MALL systems design, where words, pictures, and explanations should be near each other on the screen on mobile devices (Cakmak, 2019). Mayer and Fiorella (2014) state that when words and pictures are close to each other on the page or screen rather than far from each other, more learning from multimedia messages avoids the split-attention effect. Therefore, to address spatial separation, the physical distance between relevant information should be reduced by displaying relevant information near the learning content. However, there are some contradictions regarding the results of the spatial continuity principle. For example, Schroeder and Cenkci (2018) noted that spatial continuity designs reduce extraneous cognitive load and improve learning retention. Nevertheless, Florax and Ploetzner (2010) did not find any benefits of the spatial continuity principle. Further, motivation and attitudes toward multimedia learning are important. For example, Cabero-Almenara et al. (2019) and Wojciechowski and Cellary (2013) used technology acceptance model (TAM) to understand learners' attitudes in AR learning systems. However, few approaches have discussed learning performance, cognitive load, and attitude toward technology in the same study (Wang, Mendori, & Hoel, 2019). Therefore, it is crucial to investigate the learning performance, cognitive load, and technology acceptance of the mobile AR learning system design based on the spatial continuity principle.

9.3 The AR Compound Verb Learning Systems

9.3.1 Difficulties of Japanese Compound Verb Acquisition

Japanese compound verbs are similar to English phrasal verbs, which are formed by combining two verbs. The meaning of a compound verb is built by the interaction between the meanings of these two single verbs. For example, the compound verb tori-ireru consists of a single verb toru and ireru. In the compound verb tori-ireru, toru is called V1 (Verb1), and ireru is called V2 (Verb2). Compound verbs often appear in everyday conversations in Japanese but are not easy to learn for Japanese learners (Matsuda, 2000, 2002). Specifically, there are two main difficulties in compound verb acquisition: (1) the lack of clarity in verb combinations and (2) the opacity of meanings. In the Japanese language, not all two verbs can form a compound verb, and there are no specific rules applied to verb combinations (Chen, 2007). It is difficult to confirm which two single verbs form a compound verb and the correct order of combination (i.e., V1-V2 or V2-V1). Therefore, it is important to facilitate learners in improving the clarity of verbs. The opacity of the meaning of compound verbs is caused by the polysemous characteristic of compound verbs and the difficulty in distinguishing their meanings and V1 or V2 word meanings (Matsuda, 2002). These two difficulties make it challenging to learn compound verbs, and learners have struggled to understand and retain the meaning of the verbs (Chen, 2007; Sano, 2004). Hence, we designed and developed two AR learning systems to facilitate the acquisition of Japanese compound verbs to address these difficulties (Geng & Yamada, 2019, 2020a, 2020b).

9.3.2 AR Compound Verb Learning Systems

The AR compound verb learning systems were based on marker-based AR and developed with Unity and Vuforia. These systems can be installed on iOS and Android mobile devices. In these systems, learners are shown the meaning of verbs using 3D animations. The systems have 11 cards printed with the Japanese characters of the single verb as markers and can be combined to form eight compound verbs. Concerning the meanings of single verbs, after the learner scans the single

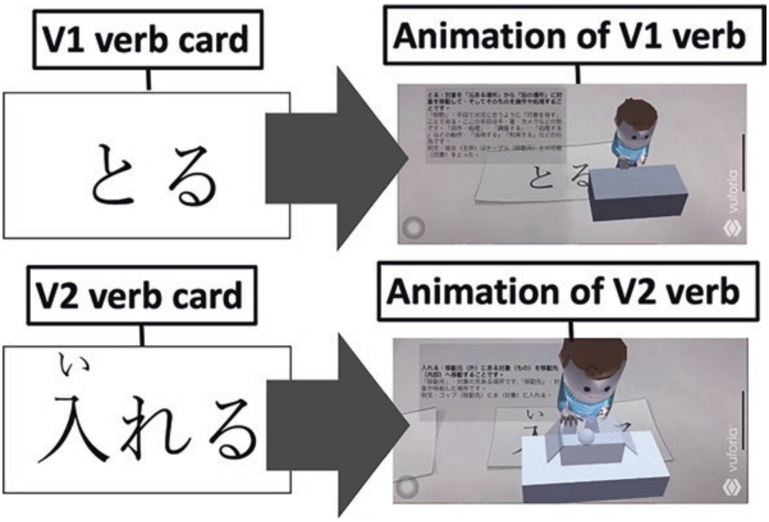


Fig. 9.1 Cards and animations of single verbs V1 and V2

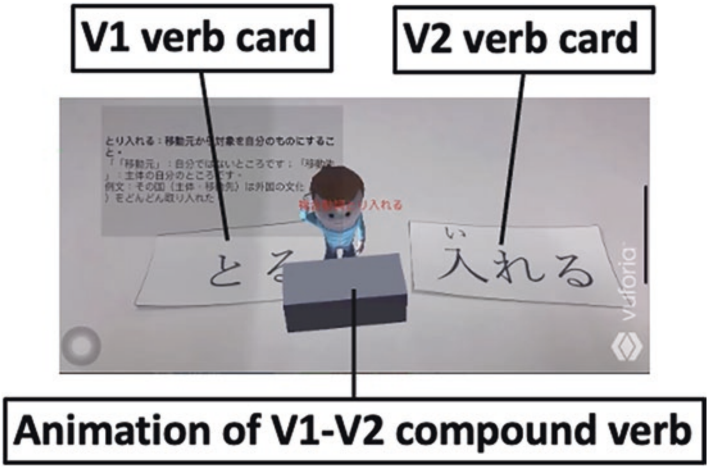


Fig. 9.2 Animations of compound verb V1-V2

verb card, an animation of the verb meaning will be displayed on the card through the smartphone screen (see Fig. 9.1). As shown in Fig. 9.2, the meaning of compound verbs is learned by combining the two verb cards to form a compound verb. After scanning the correct combination of the two verb cards, the compound verb's animation is displayed on the screen and continues until the combined card is moved out of the camera. Learners scan the verb cards to learn a single verb and then correctly combine two verb cards to learn a compound verb formed from two single verbs. The verbs (single and compound verbs) animations were created using Maya based on the verbs' image schemas. The animations were also assessed by five native Japanese speakers (including two Japanese teachers) to confirm the validity of the semantic explanations of the animations (Geng & Yamada, 2020a).

The combination judgment function of the AR learning system was designed to address the difficulties of compound verb acquisition. When the learner combines two cards to form a single verb V1 and V2, the system confirms whether the compound verb is correct. If the combination is incorrect, the system pops the message "The combination of the compound verbs is incorrect." Regarding the opacity of meaning, learning the meaning of compound verbs by single verb card combinations can aid learners in understanding the semantic distinction between single and compound verbs (Fig. 9.3).

The meanings of verbs are not only learned using animations, but verb explanations and example sentences are also used to supplement verb learning in the system. Therefore, we developed an Explanation Integration (EI) system based on the spatial continuity principle to display verb explanations and animations on the same screen. When the learner scans the card correctly or combines the verb cards to form a compound verb, the explanation and animation of the compound verb meaning are displayed on the screen (see Fig. 9.2). In addition, we developed an Explanation Separation (ES) system. In the ES system, the animation is displayed on the screen of the smartphone, while the verb explanation is provided to the learner as a paper-based learning material. Owing to the spatial separation between verb explanation and animation, there is a spatial split attention effect in the ES system. It is difficult for learners to watch the animation and read the learning material at the same time. The content of the verb explanations was the same in the two systems.

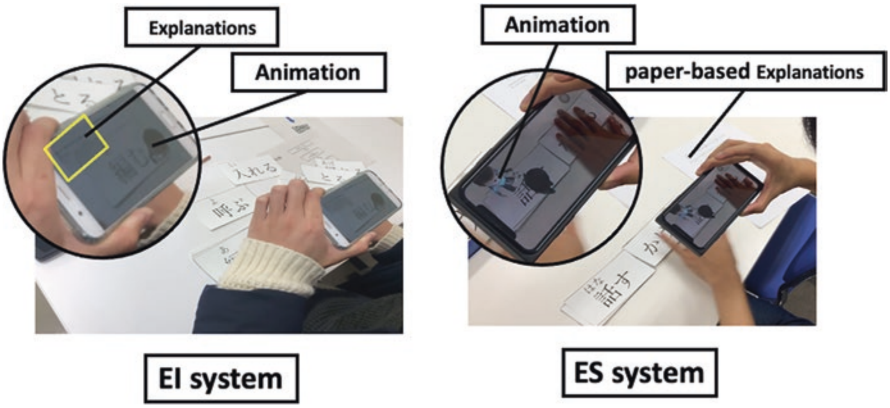


Fig. 9.3 The EI system and ES system

9.4 Research Questions

In our previous study (Geng & Yamada, 2020a, 2020b), we compared AR learning systems and paper-based images on learning performance and cognitive load. It was found that the AR learning system was more effective for knowledge retention. However, the effect of the spatial continuity principle in the design of this AR system has not been investigated. This study evaluated AR learning system designs using the spatial continuity principle by comparing the learning performance, cognitive load, and technology acceptance of EI and ES systems. Therefore, we posed the following three research questions:

Research Question 1: Do EI and ES systems facilitate the learning of Japanese compound verbs? What are the differences between the learning performances of EI and ES systems?

Research Question 2: What are the differences between the perceived cognitive loads of the EI and ES systems?

Research Question 3: What are the differences between the technology acceptance of EI and ES systems?

9.5 Methodology

9.5.1 Participants

This study was conducted in two experiments. Twenty three participants (12 males and 11 females) took part in the study. Ten participants were enrolled in Experiment 1, and 13 participants were enrolled in Experiment 2. Participants were non-native Japanese speakers from China, Vietnam, Korea, Thailand, and other countries. All participants were between the ages of 18 and 30 and had the N3 level Japanese-Language Proficiency Test or higher, which meant they could understand the descriptions and explanations of the verbs provided in the experiments.

9.5.2 The Experimental Procedure

To investigate the effectiveness of the spatial continuity principle, we conducted two experiments. Experiment 1 was implemented first, and Experiment 2 was conducted 6 months after Experiment 1. The procedure for the two experiments is shown in Fig. 9.4. The procedures of both experiments were basically the same, both lasting 90 minutes, differing only in the system used for the learning activities.

First, all participants were examined prior to the learning activities to assess their knowledge of compound verbs using a pretest. Next, all participants

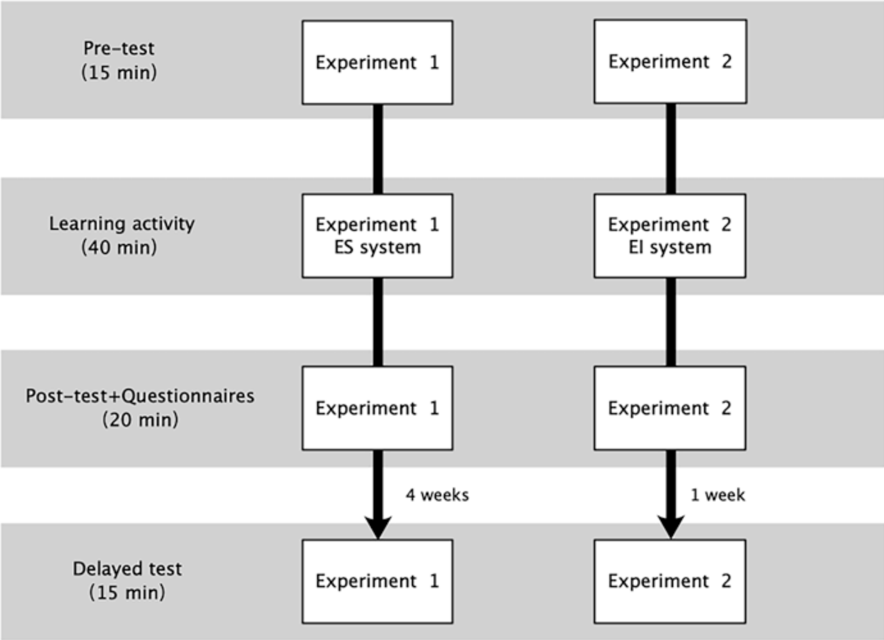


Fig. 9.4 The procedure of the two experiments

performed a 40-minute learning activity designed to facilitate the learning of Japanese compound verbs. During the learning activity, both experiment participants learned the same content, i.e., 8 compound verbs as well as 11 single verbs. Participants were free to learn Japanese verbs at their own pace. It was noted that participants in Experiment 1 used paper-based learning materials and the ES system to learn, while participants in Experiment 2 used the EI system without any paper-based learning materials during the learning activity. The posttest, cognitive load questionnaire, and technology acceptance questionnaire were administered immediately after the learning activity. In Experiment 1, the delayed test was administered 4 weeks after the posttest. However, the interval between the delayed test and the posttest was only 1 week in Experiment 2 due to COVID-19 restrictions.

9.5.3 Measuring Tools

The tests to evaluate the knowledge of learning contents, as well as the cognitive load questionnaire and the technology acceptance questionnaire, comprised the measuring tools in this study. Concerning the tests, the performance of the Japanese language single verbs and compound verbs was analyzed using the

Part 1. True or False Questions. Judge the correctness of the verbs, fill ○ in the ( ) if it is correct, and fill × in ( ) if it is incorrect.

かけ呼ぶ （ ） とり囲む（ ） 外しとる（ ）
kake-yobu tori-kakomu hazusi-toru

Part 2. Multiple-choice Questions. Choose the most correct one.
1.The questions of the single verbs.
かばんに教科書を（ ）。
A.入れる B. 呼ぶ C.かける D.とる
ireru yobu kakeru toru
2. The questions of the compound verbs.
お母さんは娘の好きな犬の絵をセーターに（ ）。
A.とり入れた B.話しかけた C.編みこんだ D.暴れこんだ
tori-ireta hanashi-kekata ami-konda abare-konda
3. The questions for distinguishing between the single verbs and the compound verbs.
子どもは、新たな知識を（ ）。
A.揃える B.とり揃える C.入れる D.とり入れる
soroeru tori-soroeru ireru tori-ireru

Fig. 9.5 Example questions of Part 1 and Part 2 on the tests

pretest, posttest, and delayed tests. For the two difficulties of compound verb acquisition, the test questions were designed and divided into two parts, and the tests were reviewed and verified by Japanese teachers with more than 10 years of teaching experience. Part 1 consisted of 15 true or false questions, and Part 2 consisted of 26 multiple-choice questions. The total number of questions was 41, and the total score was 41. As shown in Fig. 9.5, the questions in Part 1 required participants to determine whether the compound verbs appearing in the questions were present and correct in the Japanese language in order to examine the clarity of verb combinations. In Part 2, three types of questions were designed: questions of single verbs, questions of compound verbs, and questions for distinguishing between single verbs and compound verbs. The questions in Part 2 contributed to locating the opaqueness of the meaning of the compound verb. The questions used in the three tests were identical. However, the order in which questions and options appeared were different, that is, the order of questions and options were completely different for the pretest, posttest, and delayed test. Furthermore, participants were given 15 minutes to answer each test.

The cognitive load questionnaire created by Leppink, Paas, Van der Vleuten, Van Gog, and Van Merriënboer (2013) was used to measure participants' perceived cognitive loads, which are intrinsic cognitive load, extraneous cognitive load, and germane cognitive load. Using an 11-point Likert scale ranging from 0 (completely disagree) to 10 (completely agree), participants were asked to respond to ten items. There were three items each about intrinsic cognitive load and extraneous cognitive load. For each item, larger responses imply a higher cognitive load as perceived by the participants. The remaining four items were related to germane cognitive load.

The items of intrinsic cognitive load were designed to measure participants' perceptions of whether the compound verbs in the learning activity were complex. Extraneous cognitive load questions were designed to determine whether the AR compound verb learning system was clear and effective for learning the material. In addition, the questions of germane cognitive load were related to knowledge acquisition with learning activity.

The participants' attitudes in AR learning activities were evaluated using the TAM questionnaire (Cabero-Almenara et al., 2019). The TAM was developed by Davis (1989), and it consists of 15 items on 5 dimensions of acceptance of AR technology: perceived usefulness, perceived ease of use, perceived enjoyment, attitude toward its use, and intention to use. The questionnaire used a 7-point Likert scale (1 = extremely disagree to 7 = extremely agree). Specifically, four items were designed to measure perceived usefulness when using AR technology during the learning activity, three items were designed to measure the perceived ease of use, and the perceived enjoyment was three items. Moreover, attitude toward use and intention to use the AR system were three items and two items, respectively.

9.5.4 Data Analysis

To answer research question 1, the pretest, posttest, and delayed test scores were analyzed for Experiments 1 and 2. Research question 2 was analyzed by calculating the item responses of the two experiments using the cognitive load questionnaire. We also calculated the TAM questionnaire responses to analyze research question 3. The data were non-normal via normality verification and the sample size was small; hence, we conducted the analyses using nonparametric statistics.

9.6 Results

9.6.1 Research Question 1: The Learning Performance of Two AR Learning Systems

Table 9.1 presents the summary statistics of the mean, median, and standard deviation of Part 1, Part 2, and the total scores for the two experiments in the pretest, posttest, and delayed tests. Table 9.1 illustrates that the pretest total scores of Experiment 1 were higher than those of Experiment 2. A Mann-Whitney U test was conducted on the pretest of the two experiments to verify whether participants' prior knowledge of compound verbs differed. There was no statistically significant difference between the pretest scores of the two experiments from the Mann-Whitney U test results (total scores, $|Z| = 1.367$, n.s.; Part 1, $|Z| = 0.629$, n.s.; Part 2, $|Z| = 1.625$, n.s.). As can be seen from Table 9.1, in

Table 9.1 Mean, median, and SD values of the total scores

Tests		Experiment 1 (*n* = 10)			Experiment 2 (*n* = 13)		
		Mean	Median	SD	Mean	Median	SD
Pretest	Part 1	11.20	11.50	2.10	10.54	10.00	2.40
	Part 2	17.90	17.50	4.12	15.31	15.00	2.75
	Total	29.10	29.00	5.80	25.85	26.00	4.63
Posttest	Part 1	14.20	14.50	1.03	13.69	13.00	0.86
	Part 2	20.00	21.00	3.30	19.38	20.00	1.76
	Total	34.20	35.50	4.08	33.08	34.00	1.98
Delayed test	Part 1	12.80	13.00	1.14	13.15	14.00	2.08
	Part 2	19.10	18.50	4.53	20.54	21.00	3.02
	Total	31.90	32.00	5.22	33.69	34.00	4.25

Table 9.2 The Wilcoxon signed-rank tests on pre-test, post-test, and delayed test

	Tests	Experiment 1 (*n* = 10)		Experiment 2 (*n* = 13)	
		|Z|	*p*	|Z|	*p*
Part 1	Post-pretests	2.816**	0.005	2.953**	0.003
	Delayed-pretests	2.159*	0.031	2.849**	0.004
	Post-delayed tests	2.263*	0.024	1.080	0.280
Part 2	Post-pre-tests	2.106*	0.035	3.192**	0.001
	Delayed-pre tests	1.207	0.228	3.022**	0.003
	Post-delayed tests	1.259	0.208	1.298	0.194
Total	Post-pre-tests	2.677**	0.007	3.186**	0.001
	Delayed-pre tests	1.963*	0.050	3.186**	0.001
	Post-delayed tests	2.051*	0.040	1.034	0.301

†$p < 1$,*$p < 0.05$, **$p < 0.01$

Experiments 1 and 2, Part 1, Part 2, and the total scores on the post- and delayed tests were higher than the pretest. Therefore, Wilcoxon signed-rank tests were performed for the three test scores (see Table 9.2). The results revealed that both experiments' posttests were statistically higher than the pretest, and the delayed test was higher than the pretest in Part 1 scores and total scores. However, the results of the delayed test and pretest scores of Part 2 showed no statistically significant difference in Experiment 1 ($|Z| = 1.207$, n.s.) and a significant difference between the two tests in Experiment 2 ($|Z| = 3.022$, $p < 0.01$). In addition, Table 9.2 shows that no significant differences were found between the scores of the delayed test and the posttest in Experiment 2. In contrast, Part 1 scores and total scores of the delayed test in Experiment 1 were significantly lower than those of the posttest (Part 1 scores, $|Z| = 2.263$, $p < 0.05$; total scores, $|Z| = 2.051$, $p < 0.05$).

Mann-Whitney U tests were used to compare the differences in learning performance between the two experiments. The change in test scores between the two experiments was the difference between the two test scores. For example, score changes in post-pretests were the posttest scores minus the pretest scores. Table 9.3 provides the results of the Mann-Whitney U tests on the changes in the

Table 9.3 The Mann-Whitney U tests on the changes in three test scores between Experiment 1 and Experiment 2

Changes in test scores	Experiment 1 (n = 10)		Experiment 2 (n = 13)			
	Mean	SD	Mean	SD	|Z|	p
Changes in post-pretests	5.10	3.54	7.23	3.96	1.249	0.223
Changes in delayed-pretests	2.80	3.43	7.85	4.95	2.434*	0.016
Changes in post-delayed tests	2.30	2.79	−0.62	3.25	2.190*	0.031

†p < 1,*p < 0.05, **p < 0.01

Table 9.4 The Mann-Whitney U test results on cognitive load of the two experiments

Cognitive load	Experiment 1 (n = 10)		Experiment 2 (n = 13)			
	Mean	SD	Mean	SD	|Z|	p
Item 1	5.20	2.15	7.54	1.39	2.735**	0.006
Item 2	4.70	3.13	6.62	1.45	1.188	0.257
Item 3	4.10	3.35	4.77	3.00	0.562	0.605
Intrinsic cognitive load	4.67	2.23	6.31	1.54	1.926†	0.058
Item 4	2.30	3.02	2.46	1.81	0.893	0.410
Item 5	0.80	1.62	2.00	2.08	2.167*	0.036
Item 6	1.20	1.40	1.92	2.10	0.837	0.446
Extraneous cognitive load	1.43	1.48	2.13	1.89	1.068	0.300
Item 7	7.40	1.65	7.54	1.39	0.032	0.976
Item 8	7.70	2.00	7.23	1.92	0.347	0.738
Item 9	7.50	1.90	6.46	2.03	1.135	0.284
Item 10	7.90	1.60	7.85	0.99	0.129	0.927
Germane cognitive load	7.66	1.63	7.27	1.29	0.281	0.803

†p < 1,*p < 0.05, **p < 0.01

three test scores between Experiments 1 and 2. It can be seen from the data in Table 9.3 that Experiment 2 had a significantly higher change in score between the delayed test and pretest than Experiment 1 ($|Z| = 2.434$, $p < 0.05$). In addition, Experiment 2 had a significantly lower score change between the posttest and delayed test than Experiment 1 ($|Z| = 2.190$, $p < 0.05$). These results indicate that Experiment 2 had a larger difference between the delayed test and the pretest and a smaller difference between the delayed test and the posttest compared to Experiment 1.

9.6.2 Research Question 2: The Perceived Cognitive Load of Two AR Learning Systems

In this study, the cognitive load questionnaire measured learners' perceived intrinsic cognitive load, extraneous cognitive load, and germane cognitive load during the learning activities of the two experiments. The results for each item of the questionnaire and the three types of cognitive load are shown in Table 9.4. The perceived cognitive loads of Experiments 1 and 2 were compared using the Mann-Whitney U test. There were no statistically significant differences in the extraneous and germane cognitive load between the two experiments (extraneous cognitive load, $|Z| = 1.068$, n.s.; germane cognitive load, $|Z| = 0.129$, n.s.). The intrinsic cognitive load was found to be higher in Experiment 2 than in Experiment 1 ($|Z| = 1.926$, $p < 0.1$). Furthermore, the responses of item 1 ("The compound verbs covered in the activity were very complex.") in Experiment 2 were significantly higher than those in Experiment 1 ($|Z| = 2.375$, $p < 0.01$). Interestingly, regarding item 5 ("The AR learning system was in terms of learning very ineffective."), although the mean of the responses in Experiment 2 was 2, it was significantly higher than that in Experiment 1 responses ($|Z| = 2.08$, $p < 0.05$).

9.6.3 Research Question 3: The Technology Acceptance of Two AR Learning Systems

Table 9.5 shows the responses of each item and dimension on technology acceptance for Experiments 1 and 2. To compare the difference between the two experiments in terms of the attitude toward AR technology, Mann-Whitney U tests were

Table 9.5 The Mann-Whitney U test results on technology acceptance of the two experiments

	Experiment 1 ($n = 10$)		Experiment 2 ($n = 13$)			
Technology acceptance	Mean	SD	Mean	SD	$\|Z\|$	p
Item 1	5.60	1.26	5.23	0.83	0.805	0.446
Item 2	6.20	1.32	6.00	0.71	1.226	0.257
Item 3	6.10	1.20	5.69	0.95	1.285	0.232
Item 4	5.90	1.10	4.85	1.07	2.143*	0.036
Perceived usefulness	5.95	1.11	5.44	0.63	1.624	0.115
Item 5	5.30	1.42	4.54	1.45	1.271	0.232
Item 6	6.30	0.95	5.92	1.32	0.641	0.563
Item 7	5.30	1.42	5.31	1.70	0.191	0.879
Perceived ease of use	5.63	0.95	5.26	1.11	0.530	0.605
Item 8	6.20	0.92	5.85	0.90	0.915	0.410
Item 9	6.20	0.79	5.69	1.11	1.054	0.343
Item 10	5.80	1.48	5.31	1.65	0.801	0.446
Perceived enjoyment	6.07	0.87	5.62	0.93	1.161	0.257
Item 11	6.00	1.25	5.62	0.77	1.202	0.257
Item 12	5.70	1.42	5.77	0.83	0.227	0.832
Item 13	5.00	1.83	4.69	1.80	0.519	0.648
Attitude toward its use	5.57	0.98	5.36	0.83	0.596	0.563
Item 14	6.20	1.23	5.92	1.12	0.799	0.483
Item 15	6.80	0.63	6.00	1.15	2.048†	0.088
Intention to use	6.50	0.85	5.96	1.07	1.462	0.166

$^{\dagger}p < 1$, $^{*}p < 0.05$, $^{**}p < 0.01$

conducted. The statistical tests revealed that no significant differences were found between Experiments 1 and 2 in five dimensions (perceived usefulness, $|Z| = 1.624$, n.s.; perceived ease of use, $|Z| = 0.530$, n.s.; perceived enjoyment, $|Z| = 1.161$, n.s.; attitude toward its use, $|Z| = 0.596$, n.s.; intention to use, $|Z| = 1.462$ n.s.). However, there were significant differences between the two experiments on items 4 ("my performance will increase with the use of AR") and 15 ("I would like to use the AR system to learn other learning contents and subjects"). These results suggest that participants in Experiment 1 had higher confidence in the learning effectiveness of the AR system and were more willing to use other learning contents than the participants in Experiment 2.

9.7 Discussion

9.7.1 Research Question 1: The Learning Performance of Two AR Learning Systems

The statistical results of Tables 9.1 and 9.2 show that both the EI and ES systems improved the participants' learning performance of compound verbs. This finding also supports the results of previous experiments (Geng & Yamada, 2020a), demonstrating that the mobile AR learning system is effective for Japanese compound verbs. However, with regard to the acquisition of compound verb meanings, this study found that the EI system designed based on the principle of spatial continuity was more effective. This finding provides evidence of the spatial continuity principle (Mayer, 2014). Compared to the ES system, the EI system integrated the explanations of verb meanings and animation into the same screen, reducing the split attention on two sources of information. Since the ES system's animations are strongly integrated with the explanation texts, it provides the learners with active learning opportunities, allowing the information about the verbs to be retained in working memory for longer and making learning more effective (Jiang, Renandya, & Zhang, 2017). The results of the Mann-Whitney U tests in Table 9.3 further support that the EI system is more effective in retaining knowledge. In contrast to Schroeder and Cenkci (2018), the present study provides evidence for the effectiveness of dynamic interactive learning materials using spatial continuity principles. These findings may help us understand the spatial continuity principle of MALL. Unfortunately, the study differed in the time interval between the posttest and the delayed test for the two experiments of the study due to COVID-19. These findings need to be confirmed at the same time interval.

9.7.2 Research Question 2: The Perceived Cognitive Load of Two AR Learning Systems

According to the multimedia learning principle (Mayer, 2014), the spatial continuity principle can reduce the extraneous cognitive load of materials and instructions. However, the difference between the extraneous cognitive load of the ES and EI systems observed in this study was not significant. The results in Table 9.3 do not support the idea that learners have less extraneous cognitive load using mobile AR systems designed based on the spatial continuity principle. One possible explanation for this finding is that the extraneous cognitive load was low for both the EI and ES systems, with mean values of no more than 3.

Furthermore, the results of item 5 indicated that learners of the EI system perceived the AR learning system as less effective than those in the ES system, despite the contrary results for the actual learning performance. A possible explanation for these results is that the EI system lacks sufficient learning content because it does not explicitly cue all compound verbs present. However, the ES system explicitly explains the learning content of paper-based materials. This explanation might be confirmed in the statistical results of item 1 that EI system learners perceived more complex learning content than ES system learners. This finding suggests the importance of goal setting for language acquisition systems. Therefore, improvement of the EI system is needed to clarify learning goals in future investigations.

9.7.3 Research Question 3: The Technology Acceptance of Two AR Learning Systems

There was no difference in the five dimensions of attitudes toward AR technology between the participants of the two systems. This finding suggests that the MAR (mobile augmented reality) language learning system based on the spatial continuity principle did not influence attitudes toward AR technology. Furthermore, the AR technology acceptance of the EI and ES systems was higher than that of Cabero-Almenara et al. (2019). Specifically, all the means of the responses on the five dimensions were higher than 5, which indicates a high level of acceptance of the mobile AR language learning system. However, the results of items 4 and 15 revealed that the ES system learners had higher confidence in the AR system's learning effectiveness and were willing to use other AR systems to learn content. This was similar to the result of the cognitive load questionnaire item 4, which might also be caused by the lack of clear guidance on learning goals. These results need to be further explored, as our study involved a small sample size.

In general, therefore, the findings of this research provide insights for supporting the spatial continuity principle (Mayer & Fiorella, 2014), which promotes knowledge retention when text and explanations are close to each other. The principal theoretical implication of this study is that it validated the effectiveness of the spatial continuity principle for dynamic 3D animation and narration in MAR learning systems, as most previous studies have been based on 2D or static 3D pictures. Therefore, this study provides new insight and opportunities for educators and instructional designers to apply new technologies to the MALL. It implies that the MAR system allows the use of the principle of spatial continuity in the design and development of instruction to increase the learning outcomes. The cognitive theory of multimedia learning has three assumptions: dual channel, limited capacity, and active processing. In other words, information from the channels is processed by organizing them into coherent mental representations that are integrated into previously acquired knowledge (Santos et al., 2013). The MAR learning system with 3D animation and textual narration designed according to the principle of spatial continuity provides the opportunity to explain the interrelationship between the conceptual structure of information and allows the learner to build a coherent mental model. The 3D animation allows further processing of images in the visual channel of the learner to form causal chains of images (Mayer & Moreno, 2002), facilitating the integration process with textual information and the results of prior knowledge. Also, the design of the spatial continuity principle reduces the distance between 3D animations and textual narration, and thus the switching between relevant representations becomes efficient, allowing learners to construct accurate mental models in real time (Ganier, 2004). Furthermore, the findings of this study also may well have a bearing on the design of explicit learning goals. The current data highlight the importance of learning goals design in multimedia learning. Since the unclear learning goals design might lead to an increase in extraneous load, thus making the learners perceive less effective learning.

9.8 Conclusion and Future Work

This study compared the learning performance, cognitive load, and attitudes toward AR technology of two mobile AR learning systems and investigated the spatial continuity principles of AR learning systems. Learners using the EI system were more effective than learners using the ES system in terms of meaning and knowledge retention of compound verbs, as evidenced by the test and questionnaire data. The results indicated a high level of acceptance of the MAR language learning system. However, we found no differences in the extraneous cognitive load and attitudes toward AR technology between the EI and ES systems. This might be because systems using the spatial continuity principle did not provide explicit learning goals to learners.

This study has several limitations. First, the cognitive load was measured using other methods, rather than questionnaires, such as behavioral assessments. Second, the sample size of this study was too small. Furthermore, we did not control for the same time interval of the tests. Additional studies are needed to explore the learning process of the mobile AR learning system using the principle of spatial continuity. For example, the relationship between learning processes and learning performance is investigated using learning analytics (e.g., Geng & Yamada, 2020b; Kaneko, Saito, Nohara, Kudo, & Yamada, 2018).

Acknowledgments This research was supported by JSPS kaken-hi (Grant No. JP19H01716, JP20H01727).

References

Akçayır, M., & Akçayır, G. (2017). Advantages and challenges associated with augmented reality for education: A systematic review of the literature. *Educational Research Review, 20*, 1–11. https://doi.org/10.1016/j.edurev.2016.11.002

Azuma, R. T. (1997). A survey of augmented reality. *Presence: Teleoperators and Virtual Environments, 6*(4), 355–385. https://doi.org/10.1162/pres.1997.6.4.355

Boonbrahm, S., Kaewrat, C., & Boonbrahm, P. (2015). Using augmented reality technology in assisting English learning for primary school students. In *Lecture Notes in Computer Science International Conference on Learning and Collaboration Technologies* (pp. 24–32). https://doi.org/10.1007/978-3-319-20609-7_3.

Cabero-Almenara, J., Fernández-Batanero, J. M., & Barroso-Osuna, J. (2019). Adoption of augmented reality technology by university students. *Heliyon, 5*(5), e01597. https://doi.org/10.1016/j.heliyon.2019.e01597

Cakmak, F. (2019). Mobile learning and mobile assisted language learning in focus. *Language and Technology, 1*(1), 30–48.

Chen, X. (2007). Nihongo fukugodosi no syutokujyoukyou to sido hei no mondaiteiki (the acquisition of Japanese compound verbs and implications for teaching: An examination of 'Au' and 'komu' at Xi'an Foreign Language university). *Forum of International Development Studies, 35*(1), 93–102.

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly, 13*(3), 319–340. https://doi.org/10.2307/249008

Florax, M., & Ploetzner, R. (2010). What contributes to the split-attention effect? The role of text segmentation, picture labelling, and spatial proximity. *Learning and Instruction, 20*(3), 216–224. https://doi.org/10.1016/j.learninstruc.2009.02.021

Ganier, F. (2004). Factors affecting the processing of procedural instructions: Implications for document design. *IEEE Transactions on Professional Communication, 47*(1), 15–26. https://doi.org/10.1109/tpc.2004.824289.

Geng, X., & Yamada, M. (2019). Development and design of a compound verb AR learning system employing image schemas. In *15th International Conference on Mobile Learning, 2019. ML2019* (pp. 73–80).

Geng, X., & Yamada, M. (2020a). An augmented reality learning system for Japanese compound verbs: Study of learning performance and cognitive load. *Smart Learning Environments, 7*(1), 1–19.

Geng, X., & Yamada, M. (2020b). The development and evaluation of an augmented reality learning system for Japanese compound verbs using learning analytics. In *IEEE International Conference on Teaching, Assessment, and Learning for Engineering, 2020* (pp. 71–76).

Hsu, T. C. (2017). Learning English with augmented reality: Do learning styles matter? *Computers and Education, 106*, 137–149. https://doi.org/10.1016/j.compedu.2016.12.007

Ibáñez, M. B., Di Serio, A., Villarán, D., & Delgado-Kloos, C. (2016). The acceptance of learning augmented reality environments: A case study. In *IEEE 16th International Conference on Advanced Learning Technologies (ICALT), Austin, United States, 2016* (pp. 307–311). Retrieved from https://doi.ieeecomputersociety.org/10.1109/ICALT.2016.124

Ibrahim, A., Huynh, B., Downey, J., Hollerer, T., Chun, D., & O'donovan, J. (2018). Arbis pictus: A study of vocabulary learning with augmented reality. *IEEE Transactions on Visualization and Computer Graphics, 24*(11), 2867–2874. https://doi.org/10.1109/TVCG.2018.2868568

Jiang, D., Renandya, W. A., & Zhang, L. J. (2017). Evaluating ELT multimedia courseware from the perspective of cognitive theory of multimedia learning. *Computer Assisted Language Learning, 30*(7), 726–744. https://doi.org/10.1080/09588221.2017.1359187

Kaneko, K., Saito, Y., Nohara, Y., Kudo, E., & Yamada, M. (2018). Does physical activity enhance learning performance?: Learning effectiveness of game-based experiential learning for university library instruction. *The Journal of Academic Librarianship, 44*(5), 569–581. https://doi.org/10.1016/j.acalib.2018.06.002

Leppink, J., Paas, F., Van der Vleuten, C. P., Van Gog, T., & Van Merriënboer, J. J. (2013). Development of an instrument for measuring different types of cognitive load. *Behavior Research Methods, 45*(4), 1058–1072. https://doi.org/10.3758/s13428-013-0334-1

Matsuda, F. (2000). Fukuguodoushi no Imi rikai houryaku no jittai to syutoku konnanten (Acquisition of compound verbs in Japanese by advanced learners of Japanese: Learners' strategies of comprehending the meanings of compound verbs). *Japanese Language Education, 20*(1), 52–65.

Matsuda, F. (2002). Nihongogakusyusya ni yoru fukugoudoushi "komu" no syutoku [Acquisition of compound verb "komu" by Japanese learners]. *Japanese Language Education in the World, 12*(1), 43–62.

Mayer, R. E. (2009). *Multimedia learning* (2nd ed.). Cambridge University Press.

Mayer, R. E. (2014). *The Cambridge handbook of multimedia learning* (2nd ed.). Cambridge University Press.

Mayer, R. E., & Fiorella, L. (2014). Principles for reducing extraneous processing in multimedia learning: Coherence, signaling, spatial contiguity, and temporal contiguity principles. *The Cambridge handbook of multimedia learning* (pp. 27–42). Cambridge University Press. https://doi.org/10.1017/CBO9781139547369.015.

Mayer, R. E., & Moreno, R. (2002). Animation as an aid to multimedia learning. *Educational Psychology Review, 14*(1), 87–99. https://doi.org/10.1023/A:1013184611077

Nincarean, D., Alia, M. B., Halim, N. D. A., & Rahman, M. H. A. (2013). Mobile augmented reality: The potential for education. *Procedia—Social and Behavioral Sciences, 103*, 657–664. https://doi.org/10.1016/j.sbspro.2013.10.385

Sano, K. (2004). JFL Gakusyusya no fukugoudoushi no jittai to konnanten: Annke-to cyosahoukoku (the actual state and difficulty of acquiring compound verbs by JFL learners: A questionnaire survey report). *Language Culture and Japanese Language Education, 29*(1), 88–91.

Santos, M. E. C., Chen, A., Taketomi, T., Yamamoto, G., Miyazaki, J., & Kato, H. (2013). Augmented reality learning experiences: Survey of prototype design and evaluation. *IEEE Transactions on Learning Technologies, 7*(1), 38–56. https://doi.org/10.1109/TLT.2013.37

Santos, M. E. C., Lübke, A. I. W., Taketomi, T., Yamamoto, G., Rodrigo, M. M. T., Sandor, C., & Kato, H. (2016). Augmented reality as multimedia: The case for situated vocabulary learning. *Research and Practice in Technology Enhanced Learning, 11*(1), 4. https://doi.org/10.1186/s41039-016-0028-2

Schroeder, N. L., & Cenkci, A. T. (2018). Spatial contiguity and spatial split-attention effects in multimedia learning environments: A meta-analysis. *Educational Psychology Review, 30*(3), 679–701. https://doi.org/10.1007/s10648-018-9435-9

Sweller, J. (2010). Element interactivity and intrinsic, extraneous, and germane cognitive load. *Educational Psychology Review, 22*(2), 123–138. https://doi.org/10.1007/s10648-010-9128-5

Walczak, K., Wojciechowski, R., & Cellary, W. (2006). Dynamic interactive VR network services for education. In *Proceedings of the ACM Symposium on Virtual Reality Software and Technology* (pp. 277–286).

Wang, J., Mendori, T., & Hoel, T. (2019). Strategies for multimedia learning object recommendation in a language learning support system: Verbal learners vs. visual learners. *International Journal of Human–Computer Interaction, 35*(4–5), 345–355. https://doi.org/10.1080/10447318.2018.1543085

Wojciechowski, R., & Cellary, W. (2013). Evaluation of learners' attitude toward learning in ARIES augmented reality environments. *Computers and Education, 68*, 570–585. https://doi.org/10.1016/j.compedu.2013.02.014

Chapter 10
Social Robots in Education: Conceptual Overview and Case Study of Use

Josef Guggemos, Sabine Seufert, Stefan Sonderegger, and Michael Burkhard

10.1 Introduction

Social robots could become an essential part of the educational infrastructure (Belpaeme, Kennedy, Ramachandran, Scassellati, & Tanaka, 2018; Cheng, Sun, & Chen, 2018; Papadopoulos et al., 2020). A social robot can be defined as "an autonomous or semi-autonomous robot that interacts and communicates with humans by following the behavioral norms expected by the people with whom the robot is intended to interact" (Bartneck & Forlizzi, 2004, p. 592). The quasi-standard robot in educational settings is currently Nao (see Fig. 10.1), developed by SoftBank Robotics (Belpaeme et al., 2018). In recent years, Pepper (Fig. 10.1), also a humanoid robot from SoftBank Robotics, has been increasingly used (Woo, LeTendre, Pham-Shouse, & Xiong, 2021).

Social robots can be used to teach about robots or as teaching aids (Guggemos & Seufert, 2021; Mubin, Stevens, Shahid, Mahmud, & Dong, 2013). When teaching about social robots, they are the actual content of instruction, for example, to teach computational thinking (Ching, Hsu, & Baldwin, 2018; Guggemos, 2021), or they are used to evoke interest in technology. This latter kind of use will not be a part of this chapter. Rather, how social robots (in collaboration with teachers) can carry out selected duties in the classroom will be addressed. In this vein, frequently mentioned roles are (Belpaeme et al., 2018; Mubin et al., 2013; Woo et al., 2021) teaching assistant, tutor, and peer. Social robots as teaching assistants have gained much attention, especially in language learning (van den Berghe, Verhagen, Oudgenoeg-Paz, van der Ven, & Leseman, 2019). For instance, Alemi, Meghdari, and Ghazisaedy

J. Guggemos (✉) · S. Seufert · S. Sonderegger · M. Burkhard
Institute for Educational Management and Technologies, University of St. Gallen, St. Gallen, Switzerland
e-mail: josef.guggemos@unisg.ch

D. Ifenthaler et al. (eds.), *Orchestration of Learning Environments in the Digital World*, Cognition and Exploratory Learning in the Digital Age,
https://doi.org/10.1007/978-3-030-90944-4_10

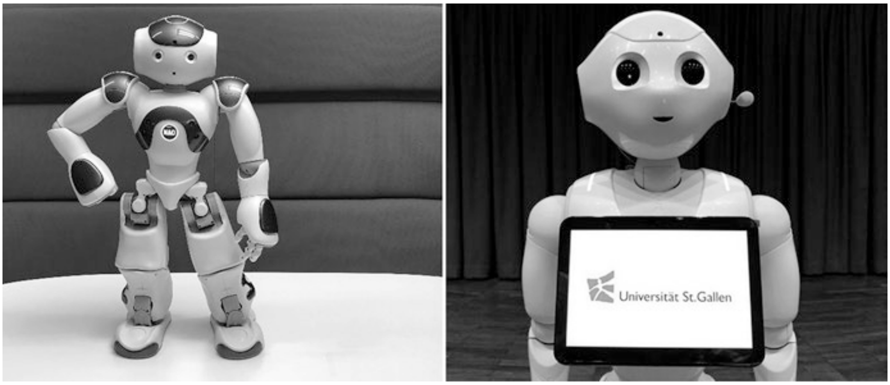


Fig. 10.1 On the left, the social robot Nao (height = 58 cm); on the right, Pepper (height = 120 cm)

(2015) used Nao as a teaching assistant for English as a foreign language among 12–13-year-olds. Nao played games with students, called them out, performed songs, positively reinforced correct answers, and made mistakes on purpose. Nao has also been successfully used to assist teachers in storytelling for kindergarten children (Conti, Cirasa, Di Nuovo, & Di Nuovo, 2020). Besides assisting the teacher, social robots can also act as tutors or peers. The role of a peer might have the advantage that students are more willing to accept mistakes from a peer in comparison to a tutor (Baxter, Ashurst, Read, Kennedy, & Belpaeme, 2017). In a long-term experimental study, Vogt et al. (2019) used Nao as an English foreign language tutor with 6-year-old children. The children played educational games on a tablet; Nao provided verbal support in the form of instructions, translations, and feedback and nonverbal support in the form of gestures. Concerning social robots as peers, Jamet, Masson, Jacquet, Stilgenbauer, and Baratgin (2018) reported on the overall positive effects of a learning by teaching approach, based on a review of the literature. Hood, Lemaignan, and Dillenbourg (2015), for example, asked 7–8-year-olds to teach a Nao robot handwriting: the robot writes a letter and asks for feedback; the child provides feedback via demonstration using a digital pen and a tablet; the robot responds to the feedback until the child is satisfied with the robot's performance.

Overall, social robots show promising results in terms of cognitive and affective learning outcomes (Belpaeme et al., 2018). However, a physical presence is not imperative for social agents because they do not have to carry out physical tasks, like industrial robots. Hence, social agents could also appear virtually on a screen as a telepresence robot or as a virtual agent (Li, 2015). Since physical presence incurs additional cost, including servicing and transporting the robot to the venue, this type of usage has to be justified (Belpaeme et al., 2018). The literature review of Li (2015) concludes that people respond more favorably to physical robots in comparison to virtually present robots. In light of this, social robots may actually have an added value in comparison to virtual agents, due to their physical presence.

This chapter aims at shedding light on the phenomenon of social robots in education. To this end, Section 10.2 characterizes social robots by means of visual appearance and social capabilities, as well as autonomy and intelligence. Section 10.3 describes how scenarios for the use of social robots can be identified, including ethical questions that need to be addressed and how technology acceptance of such robots can be evaluated. Section 10.4 explains how Lexi, a Pepper-model type, was used as a teaching assistant on an academic writing course and reports findings from this project. Section 10.5 outlines the conclusions and provides avenues for further research.

10.2 Characteristics of Social Robots

Three characteristics may be useful to describe social robots (Baraka, Alves-Oliveira, & Ribeiro, 2020): visual appearance, social capabilities, and autonomy and intelligence.

10.2.1 Visual Appearance

In education, bio-inspired social robots—humanoid and zoomorphic—seem to be prevalent (Baraka et al., 2020). Figure 10.2 shows various types of regularly used robots, beyond Nao and Pepper (Fig. 10.1). Humanoid robots resemble the human body in varying degrees. Androids, a subset of humanoid robots, are designed to be highly anthropomorphic. A special type of android is a geminoid, which duplicates an existing person (Nishio, Ishiguro, & Hagit, 2007). The relationship between humanlike appearance and human affinity toward social robots may not be linear. Rather, Mori posited the idea of an "uncanny valley" (Mori, MacDorman, & Kageki, 2012): human affinity toward social robots increases with humanlike appearance until the uncanny valley is reached whereby people experience an eerie sensation. In 48% of the studies reviewed by Belpaeme et al. (2018), a Nao model is used; further humanoid robots are Wakamaru (5%), Robovie (4%), and Bandit (4%). Animal-shaped social robots can fall into two categories (Baraka et al., 2020):

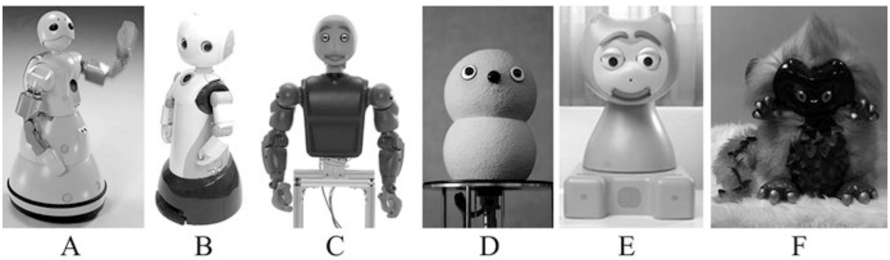


Fig. 10.2 Examples of social robots in education: (**a**) Wakamaru, (**b**) Robovie, (**c**) Bandit, (**d**) Keepon, (**e**) iCat, and (**f**) DragonBot

familiar versus unfamiliar and real as opposed to imaginary. The iCat (4%) resembles a cat and is an example of a real and familiar animal. The DragonBot (4%) is an example of a familiar imaginary animal, and the Keepon (6%) an example of an unfamiliar imaginary animal. These percentages also refer to the studies reviewed by Belpaeme et al. (2018).

Evidence is available concerning the influence of the robot's visual appearance on desired outcomes (see also Sect. 10.4). The literature review of Mou, Shi, Shen, and Xu (2020) demonstrated a substantial influence of visual appearance on perceived robot personality. Interestingly, humanoid robots may not necessarily be superior to animal-shaped robots: people may form unrealistic expectations that cannot be met at the current technological state of the art and are eventually disappointed (Henschel, Laban, & Cross, 2021). In the context of education, however, studies that compare different types of robots in the same setting are scarce (Belpaeme et al., 2018).

10.2.2 Social Capabilities

Social capabilities address the ways in which social robots engage in interactions with humans. An important aspect is verbal and nonverbal communication (Mavridis, 2015). Social robots can communicate using a combination of natural speech, motion, lights, and sounds (Baraka et al., 2020). As can be seen from Figs. 10.1 and 10.2, all types of robots have eyes. Gaze is a crucial element in nonverbal communication (Admoni & Scassellati, 2017), and eye contact is important in social encounters (Ahmad, Mubin, & Orlando, 2017). Niculescu, van Dijk, Nijholt, Li, and See (2013) provided evidence for the importance of voice characteristics and language cues for the perceived quality of interaction with the robot. Salem, Kopp, Wachsmuth, Rohlfing, and Joublin (2012) described the positive influence of social robot gestures on its evaluation by users. All these features seem to influence the perceived robot personality (Mou et al., 2020), which can positively impact the willingness to interact with a robot (Fong, Nourbakhsh, & Dautenhahn, 2003).

Empathy might be a social capability of specific importance in human-robot interaction (Baraka et al., 2020). Leite, Castellano, Pereira, Martinho, and Paiva (2014, pp. 330–331) presented a model demonstrating how social robots can show empathy. First, the affective state of the user is identified using visual and acoustical cues, as well as information about the current situation. Identified emotions could be anger, contempt, disgust, fear, happiness, neutral, sadness, and surprise. Moreover, the situation could offer information. For instance, if a student has just received inconvenient feedback, negative emotions may be likely. Second, based on the user's current affective state, the robot generates an empathic response, for example, an appropriate facial expression. Third, if the affective state is negative, the robot takes action to reduce the distress of the user. To this end, it shows supportive behaviors: "information support" (advice or guidance), "tangible assistance" (concrete support, e.g., by providing services), "esteem support" (reinforcing the user's sense of competence), and "emotional support" (an expression of caring or connectedness). Fourth, since remembering past interactions is crucial for building relationships, the robot utilizes information from previous interactions with the user to generate a dialogue that aims to give the user the feeling of "being cared for."

10.2.3 Autonomy and Intelligence

10.2.3.1 Autonomy

Autonomy can be defined as "the extent to which a robot can **sense** its environment, **plan** based on that environment, and **act** upon that environment with the intent of reaching some **task-specific goal** (either given to or created by the robot) without external **control**" (Beer, Fisk, & Rogers, 2014, p. 77). Table 10.1 presents a taxonomy that outlines the autonomy continuum, ranging from fully teleoperated to fully autonomous, depending on the level of human intervention. According to the definition of autonomy, it can be characterized by the involvement of the robot in sensing, planning, and acting. However, this taxonomy of social robot autonomy neither implies that full autonomy in educational settings is possible at the current technological state of the art nor that it is desirable. Rather, the level of autonomy should be a design choice (Baraka et al., 2020). It may be helpful in the design process of cases of use as the taxonomy can split up tasks and assign (sub-)tasks to either the robot or the teacher. Table 10.1 depicts the continuum of robot autonomy and illustrates it using the example of classroom management.

At the current technological state of the art, autonomy—even at a low level—is hard to achieve. Woo et al. (2021) reviewed studies in naturalistic classroom settings ("in the wild"). They found that in only 23 out of 126 studies (18%) the robot acted autonomously, at least to some degree.

For research purposes, the Wizard of Oz technique has regularly been used; in other words, this involves "a person (usually the experimenter, or a confederate) remotely operating a robot, controlling any of a number of things, such as its movement, navigation, speech, gestures, etc." (Riek, 2012, p. 119). By means of this, a desired level of autonomy can be simulated. Guidelines on how to conduct studies using the Wizard of Oz technique are available (Riek, 2012).

10.2.3.2 Intelligence

Educational situations are characterized by a high degree of complexity and events that are difficult to predict. Hence, intelligence is necessary in order to achieve robot autonomy. Intelligence can be regarded as the "capacity of an information-processing system to adapt to its environment while operating with insufficient knowledge and resources" (Wang, 2019, p. 17). In line with the definition of

Table 10.1 Social robot autonomy in education

Level of autonomy	Description	Example from classroom management
(Assisted) teleoperation	"The robot assists the human with action implementation. However, sensing and planning is allocated to the human"	The teacher monitors the classroom to detect undesired behavior, decides that an intervention is necessary for a specific student, and prompts the robot to call the student to order
Batch processing	"Both the human and robot monitor and sense the environment. The human, however, determines the goals and plans of the task. The robot then implements the task"	Teacher and robot monitor the classroom to detect undesired behavior. The teacher utilizes the information provided by the robot to decide what to do and prompts the robot to carry out this action
Decision support	"Both the human and robot sense the environment and generate a task plan. However, the human chooses the task plan and commands the robot to implement actions"	Teacher and robot monitor the classroom to detect undesired behavior. The robot suggests potential actions. The teacher decides on the action to be carried out and prompts the robot to do so
Shared control with human initiative	"The robot autonomously senses the environment, develops plans and goals, and implements actions. However, the human monitors the robot's progress and may intervene and influence the robot with new goals and plans if the robot is having difficulty"	The robot monitors the classroom to detect undesired behavior, decides on adequate means, and carries them out. The teacher monitors the robot and provides corrective feedback to the robot
Shared control with robot initiative	"The robot performs all aspects of the task (sense, plan, act). If the robot encounters difficulty, it can prompt the human for assistance in setting new goals and plans"	The robot monitors the classroom to detect undesired behavior, decides on adequate means, and carries them out. The robot asks the teacher for help if necessary, e.g., if the classification probability does not meet a desired level
Executive control	"The human may give an abstract high-level goal […]. The robot autonomously senses environment, sets the plan, and implements action"	The teacher sets the high-level goal, e.g., optimal classroom management, and the robot uses its capabilities to carry out this complex task
Supervisory control	"The robot performs all aspects of the task, but the human continuously monitors the robot, environment, and task. The human has override capability and may set a new goal and plan"	The robot carries out all classroom management activities. During this process, the teacher might prompt the robot to act less harshly
Full autonomy	"The robot performs all aspects of a task autonomously without human intervention in sensing, planning, or implementing action"	The robot carries out all classroom management activities without any intervention from the teacher

Note: Levels of autonomy and description taken from Beer et al. (2014, p. 87) autonomy, sensing, planning, and action have to be considered. To sense the environment, social robots can rely on cameras, microphones, and bumpers, as well as tactile, 3D, sonar, and laser sensors (Pandey & Gelin, 2018). By means of speech-to-text engines, transcripts of verbal input can be obtained for further analysis. In the next step, the obtained data has to be interpreted; for example, an answer to a question needs to be classified as incorrect, or the student has to be assigned a value for happiness based on a taken picture. In the planning phase, the robot decides what action should be carried out based on the specified goals, possible actions, and available information. The available information from the robot's sensors can be complemented with further data, for example, from the learning management platform. Moreover, the robot can access information about social expectation concerning appropriate robot behavior, such as how to show empathy. Based on these elements, the robot decides what action contributes most to achieving the specified goal. Afterward, the robot carries out the selected action. This includes verbal and nonverbal reactions, for example, positive encouragement and corrective feedback via gestures, body movement, speech, light, and sound.

The robot does not have to perform the above-described processes solely by relying on its hard- and software. Rather, it can access (artificial intelligence based) remote services via WiFi and using API services. An example is emotion and sentiment analysis (see Khanal, Barroso, Lopes, Sampaio, & Filipe, 2018). Chatbots such as Jill Watson can be used to answer student questions (Goel & Polepeddi, 2016). Furthermore, if the robot should be a tutor, an intelligent tutoring system (Mousavinasab et al., 2021) could act as the basis. In this vein, it may also be beneficial for the robot to have access to an available learning analytics system (Ifenthaler, 2015).

Due to the complexity of natural classroom settings, it is important for the robot to have learning capabilities in order to improve its performance. This can be achieved by means of machine-learning methods (Mosavi & Varkonyi-Koczy, 2016). For instance, the robot can use feedback from the teacher to improve its performance of subsequent tasks (reinforcement learning: Mosavi & Varkonyi-Koczy, 2016).

10.3 Use of Social Robots

10.3.1 Task Analysis

Beer et al. (2014) argue that the level of robot autonomy is a design choice. The starting point for determining suitable levels of autonomy and identifying corresponding use scenarios may be tasks carried out by teachers. In the educational context, teaching standards (e.g., InTASC Model Core Teaching Standards; CCSSO, 2021) could be used to identify the tasks of teachers. Moreover, characteristics of high-quality learning environments could be revealing (Bransford, Brown, &

Cocking, 2000). According to Praetorius, Klieme, Herbert, and Pinger (2018), the basic characteristics of high-quality teaching are classroom management, student support, and cognitive activation. *Classroom management* addresses the fostering of desirable student behaviors while at the same time preventing undesirable ones. Examples of the former are clear rules and routines. *Student support* draws from the self-determination theory and aims at supporting experiences of competence, autonomy, and social relatedness. *Cognitive activation* deals with the involvement of students in higher-level thinking, for example, by offering tasks of suitable difficulty (Kärner, Warwas, Krannich, & Weichsler, 2021). These findings seem to be well in line with the role of social robots as teaching assistants and tutors. Overall, it might be important to start from a sound conceptual basis of how people learn and what constitutes high-quality instruction. This could prevent the development and use of such AI-based technology as an end in itself (Zawacki-Richter, Marín, Bond, & Gouverneur, 2019).

Not all teaching tasks are suitable to be performed by social robots, especially those which are critical or too complex. Critical tasks include those that have serious consequences if carried out inappropriately, for example, determining which students have to repeat the class. Ethical concerns or a lack of technology acceptance (see Sect. 3.3) could also constitute critical tasks. In terms of complexity, studies about autonomous robots reveal drivers (Woo et al., 2021). Social robots are generally used in a one-to-one or one-to-a-class setting. The robot does not have to navigate the classroom to approach individual students or teams, which is a highly complex task for a robot due to the changing environment. Moreover, a conversation within a group in a noisy environment, as can be characteristic for group work in naturalistic classrooms, causes high complexity. Besides this, the instructional content becomes increasingly complex: tutoring on the elementary level, for example, in vocabulary learning, is less complex in comparison to the university level (Handke, 2020).

However, even for complex tasks, a social robot can take over sub-tasks, for example, monitoring the classroom as part of the overall task "classroom management" (see Table 10.1). The ways in which teacher and social robot can collaborate will be outlined in the next section.

10.3.2 Task Sharing of Human and Robot

Following a symbiotic design approach, as outlined by Baraka et al. (2020, pp. 51–53), social robots and humans could collaborate in ways that benefit both parties. This is in line with the concept of hybrid intelligence (Dellermann, Ebel, Söllner, & Leimeister, 2019). Humans and smart machines have complementary capabilities that augment each other. Human strengths comprise flexibility and transfer, real empathy and creativity, annotation of arbitrary data, and common sense. Smart machines have strengths in pattern recognition, dealing with probabilities, and ensuring consistency and speed, as well as efficiency. For instance, in order to cognitively activate students, considering prior knowledge is important in order to offer a task of suitable difficulty (Sweller, 2020). The social robot could quickly make a suggestion for a suitable task based on the previous achievements of the student and their current emotional state. For the teacher, it would not be possible to access and process all the available information in a classroom setting. However, the teacher should not blindly trust in the suggestions of the robot, but critically question its decisions and actions (Dellermann et al., 2019). Eventually, the teacher could make the decision what task should be assigned to the student, also based on intuition. This final decision might be better in comparison to the decision of either the social robot or the human teacher in isolation. Furthermore, the teacher may train the social robot to improve the precision of its future predictions. For instance, the teacher can label predictions as wrong and therefore help the robot to improve its performance. Moreover, the teacher may help the robot to overcome physical obstacles and therefore reduce environmental complexity for the robot (Baraka et al., 2020). Conversely, the teacher might benefit from the social robot's feedback concerning their performance in the classroom.

The leading question to consider, however, may be: how can the teacher and social robot, as a team, best perform the tasks that are necessary to ensure a high-quality learning environment? With this in mind, Burkhard, Seufert, and Guggemos (2021a, 2021b) argue for focusing on the comparative advantages of both parties. Although a symbiotic design approach seems promising, from a conceptual point of view, teamwork with smart machines, such as social robots, is a complex endeavor and a novel research field with many open-ended questions (Seeber et al., 2020).

10.3.3 Restrictions of Social Robot Use

10.3.3.1 Ethical Aspects

Adherence to ethical standards may be necessary in order to maintain the moral legitimacy of the organization (Suchman, 1995). Sharkey (2016) and Serholt et al. (2017) discussed the ethical concerns of social robots in classrooms. Privacy is mentioned as an important aspect. To act (in part) autonomously, robots continuously evaluate their environment and collect data about students, for example, their emotional state. Privacy issues may include "amount of data; sensitivity of data; security risks such as hacking; cloud connectivity; third-party access" (Lutz, Schöttler, & Hoffmann, 2019, p. 424). A strategy has to be developed concerning how to deal with these issues. It may include transparency, student control over data, and right of access, as well as accountability and assessment (Pardo & Siemens, 2014).

Other ethical concerns are the undesired consequences of student-robot interaction. Excessive use of the robot may impede the development of social skills among students. For instance, students may form unrealistic expectations about social interactions due to the high adaptiveness and predictability of social robots; this might especially be a concern among young children (Sharkey, 2016). Moreover, it is questionable to what extent a robot should be allowed to exert power over students. Taking the example of classroom management, the robot could only be allowed to positively encourage students rather than penalize them. Furthermore, students may not always act benevolently toward robots. Brščić, Kidokoro, Suehiro, and Kanda (2015) reported abusive behavior of children, for example, punching, toward social robots operating in public places. Although such behavior may be unlikely when teachers are present, an interesting question that arises might be how will the robot be allowed to defend itself?

The role of teachers is a further ethical consideration. Concerns about the replacement of teachers by smart machines may be unfounded (Belpaeme & Tanaka, 2021; Frey & Osborne, 2017). Nevertheless, teachers need to be reassured that the intention is not to replace them by social robots (Mubin et al., 2013). Although teachers are unlikely to be replaced by smart machines, the teacher role is likely to change if teachers are expected to collaborate with social robots. Not all teachers may appreciate this new role or the input provided by the robot, for example, feedback about their own performance. Moreover, as resources are scarce, allocating more funds to social robots could indirectly affect teachers, as budgets for both them and their training might be cut.

Physical harm caused by social robots is unlikely due to various safety measures (Pandey & Gelin, 2018). However, as the robot itself cannot be held responsible, the question arises who should be held accountable for the robot's actions. If teachers should be made responsible, it needs to be clarified how they can be involved in the process of sensing, planning, and acting in order to be able to take over responsibility. However, insights into these processes may be difficult to ascertain because AI systems are often a black box. Developments in the research field of explainable AI may help to alleviate this problem (Gunning et al., 2019).

As detailed in Sect. 2.3, learning capabilities are necessary for social robots. Training opportunities in naturalistic settings are vital in order to allow the robot to learn from mistakes; not all potential circumstances can be anticipated in simulations or laboratory settings. However, to fine-tune the robot in naturalistic settings may only be acceptable if the error rates in controlled settings are markedly lower for the robot in comparison to a teacher (Vallor & Bekey, 2017). Moreover, human-generated learning data may be biased concerning, for example, gender, socioeconomic status, and migration background (Vallor & Bekey, 2017). This could reinforce prejudices because robots might be deemed to act objectively. To tackle the issue of biases in AI applications, the concept of "fair AI" may be promising (Feuerriegel, Dolata, & Schwabe, 2020).

10.3.3.2 Technology Acceptance

Eventually, over time, social robots may have to be used by both teachers and students. If teachers and students were reluctant to use social robots (in specific areas), then robots would not be useful, regardless of their actual potential value. Several frameworks are available for evaluating the acceptance of intelligent agents in education (Sohn & Kwon, 2020). Regularly used frameworks in the context of social robotics are the technology acceptance model (TAM) and the unified theory of acceptance and use of technology (UTAUT) (Fridin & Belokopytov, 2014). These models are informed by the theory of planned behavior (Scherer, Siddiq, & Tondeur, 2019). In line with this theory, predictors of the intention to use social robots are attitudes toward their specific use, subjective norms, and perceived behavioral control (Ajzen, 1991).

Social robot acceptance studies usually enrich generic technology acceptance models with characteristics of the robot, for example, the appearance or the interaction experience, such as privacy concerns. A prominent example is the study of Heerink, Kröse, Evers, and Wielinga (2010) that evaluated the acceptance of the iCat (see Fig. 10.2) by elderly users. Graaf and Allouch (2013) provide, based on a review of the literature, a comprehensive list of robot characteristics and user experience variables that are associated with the acceptance of social robots.

In terms of empirical evidence, the literature review of Naneva, Sarda Gou, Webb, and Prescott (2020) reports overall positive attitudes toward social robots. With a focus on education, Fridin and Belokopytov (2014) evaluated the acceptance of technology by teachers. Based on a sample of 18 teachers, they reported a positive association of the intention to use the Nao robot with attitudes and perceived usefulness; the overall perception of Nao was reportedly positive. However, other studies also report cautious attitudes of teachers toward social robots (Kennedy, Lemaignan, & Belpaeme, 2016; Reich-Stiebert & Eyssel, 2016). Based on a sample of 345 people, mainly university students, Reich-Stiebert and Eyssel (2015) reported a slight reluctance to engage in joint learning activities with social robots. Smakman, Konijn, Vogt, and Pankowska (2021) investigated the attitudes of important stakeholders (parents, teachers, school management, governmental policy makers, the robot industry, students) toward social robots in primary education ($N = 515$). By means of a cluster analysis, they identified five profiles: enthusiastic, practical, troubled, sceptic, and mindfully positive. As this study showed, there seems to be substantial variance in the attitudes toward social robots in education. This may be in line with the Eurobarometer 382 survey (2012), based on interviews with 26,751 citizens from the European Union. Of these, 34% named education as an area where the use of robots should be banned. Such attitudes by the public might be regarded as the social norm in the sense of the theory of planned behavior. Against this backdrop, it may be of particular importance to consider stakeholders and opinion leaders when integrating social robots into the classroom. A group of particular relevance might be parents (Smakman, Jansen, Leunen, & Konijn, 2020).

It is worth noting that technology acceptance studies usually rely on self-assessment instruments. Positive perceptions may be neither beneficial nor valid. For instance, students may perceive the robot as perfectly trustworthy and, therefore, do not rely on common sense (the comparative advantage of humans) when using the robot. Another example is perceived learning gains that do not necessarily equal actual learning outcomes (Nasir, Norman, Bruno, & Dillenbourg, 2020).

10.4 The Use of Lexi in Academic Writing

The Institute for Educational Management and Technologies at the University of St. Gallen investigated the acceptance of social robots as teaching assistants in higher education (Guggemos, Seufert, & Sonderegger, 2020). The importance of considering technology acceptance has been outlined in Sect. 3.3.2. Moreover, the empirical evidence on the use of social robots in higher education is scarce (Handke, 2018; Spolaôr & Benitti, 2017; Zhong & Xia, 2018). Existing studies primarily address their use in technical courses, for example, computer science (Abildgaard & Scharfe, 2012; Byrne, Rossi, & Doolan, 2017). From a conceptual point of view, social robots might be a valuable learning resource for students, especially in large-scale university courses (Byrne et al., 2017; Cooney & Leister, 2019). In such an environment, it is often difficult to adequately support students and answer individual questions. The use of human assistants for this purpose may not be feasible for various reasons, mainly budget constraints.

For the study, we used Lexi, a Pepper-model type (see Fig. 10.1). It can communicate verbally and nonverbally with users through speech, gestures, and facial expressions (Huang & Mutlu, 2014). In addition, Lexi can use a tablet placed on its chest to receive input and present output. Besides commercial use, for example, as a sales assistant in a shopping mall, Lexi is also used in educational settings, such as language learning among children (Tanaka et al., 2015). The setting for the present study was the introductory lecture of an academic writing course that is mandatory for all 1500 freshmen students at the University of St. Gallen. This setting is characterized by a pronounced heterogeneity in the prior knowledge of the students on the course material (Seufert & Spiroudis, 2017). Thus, this environment seems conceptually promising for the use of social robots as teaching assistants.

The unified theory of acceptance and use of technology (UTAUT) served as the theoretical framework for the study. According to this theory, the performance expectancy (PE), the effort expectancy (EE), and the social influence (SI) determine the intention of use (Venkatesh, Morris, Davis, & Davis, 2003). Sample items to capture these constructs are "Lexi could be useful for my learning success" (PE), "It would be easy for me to learn together with Lexi" (EE), and "My friends would appreciate it if I learned together with Lexi" (SI). Other constructs are only indirect predictors of the intention of use. Based on studies about the characteristics of social robots that potentially (indirectly) influence the intention of use (see Sect. 3.3.2), we selected four constructs, which are summarized in Table 10.2.

Social robots may be a specific kind of technology because they rely on AI to carry out tasks. In our case, a Microsoft Azure service was used to identify emotions based on a picture of a face. Study participants may have concerns about the handling of data collected during interactions with social robots (Lutz et al., 2019; Lutz & Tamó-Larrieux, 2020). Therefore, concerns about the use of the collected data have to be taken into account (anxiety backend). A sample item is "I would be worried about my privacy." In order to separate these concerns from possible anxieties in general interaction with social robots (Graaf & Allouch, 2013), we also captured

Table 10.2 Characteristics of the robot, definitions, and sample items

Construct	Definition	Sample item
Trustworthiness (Heerink et al., 2010)	Degree to which students perceive the robot to be competent and act with integrity	I would trust in Lexi's advice
Adaptiveness (Heerink et al., 2010)	Degree to which students believe that the social robot adapts to their (learning) needs	Lexi could adapt to my personal learning needs
Social presence (Heerink et al., 2010)	Degree to which students perceive the robot to be a social entity	Lexi appeared to me like a real person
Appearance (Pandey & Gelin, 2018)	Perception of the acoustical and visual presence of the robot	Lexi has a nice appearance

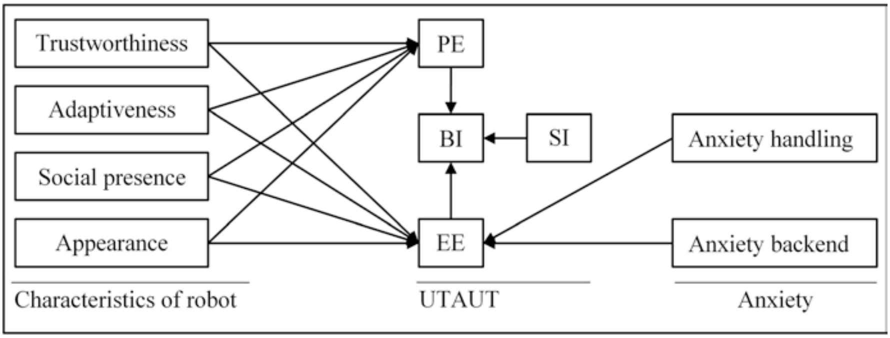


Fig. 10.3 Conceptual framework for predicting behavioral intention (BI) (Guggemos et al., 2020)

the construct "anxiety handling." A sample item is "I would be afraid of making mistakes." According to the UTAUT, anxieties negatively influence EE and thus have an (indirect) negative influence on the intention of use. Figure 10.3 summarizes the conceptual framework of the study.

In order to enable students to fairly assess the constructs presented in Table 10.2, Lexi had to provide students with a sample of its capabilities. It had to carry out activities that are representative in the context of learning at the current state of the art. For this purpose, the work of Cooney and Leister (2019) acted as the conceptual basis. Sample activities were developed to adequately represent the capabilities of the social robot (see Table 10.3). In addition to generic activities, such as greeting, typical problems and issues in academic writing were also addressed. For instance, Lexi explained how the plagiarism software worked: it guided students through the application and explained what output the algorithm of the plagiarism software generates and how the lecturer could then make a decision for each individual case. For the technical implementation, we collaborated with raumCode from Zurich, a company that specializes in social robots and AI. Lexi assisted for about 45 min during the lecture. It was connected to the projector in the venue and equipped with a headset. A video illustrates the activities according to Cooney and Leister (2019): https://unisg.link/lexi2020. After the lecture, a representative of raumCode explained

Table 10.3 Activities performed by Lexi during the course "Introduction to Academic Writing"

Activities	Tasks performed during the "Introduction to Academic Writing" course
Greeting	Introduces itself and the institute
Reading	Presents its aim: Assisting lecturers and students to foster learning Explains how plagiarism is detected by means of plagiarism software; outlines the software analyses, the outcome of the analytic process, and the decision-making process: The human is the final decider in every case (e.g., for a human-machine interaction and procedure based on complementary skills)
Alerting	Reminds the lecturer about presenting the functioning of the plagiarism software
Remote operations	Supports the lecturer by looking for sources on "greenwashing" in the database of the university's library Converses with a volunteer student: Accesses remote services during the conversation to determine the student's face characteristics, age, and mood ("happy," "surprised," "angry," "sad," and "neutral")
Clarification	Presents further material using the projector of the venue and demonstrates the functioning of the plagiarism software by giving an illustrative example
Motions	Follows the lecturer with its head; uses gestures to support its points and to express emotions

Note: Activities taken from Cooney and Leister (2019)

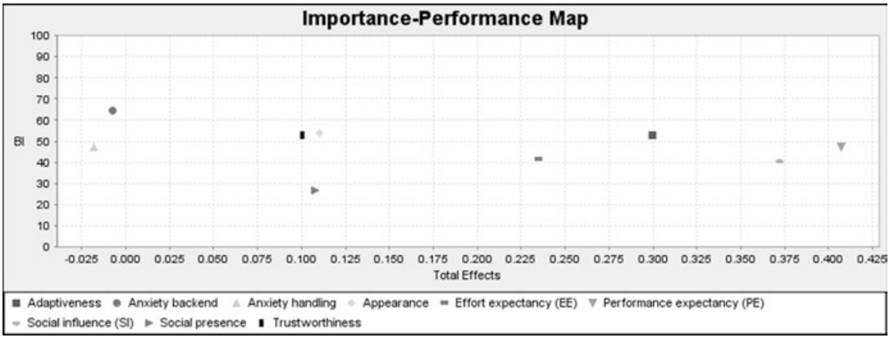


Fig. 10.4 Importance-performance map analysis (IPMA) for behavioral intention (BI) (Guggemos et al., 2020)

the functionality of the robot to the students. This may enable students to evaluate what is going on behind the scenes and, thus, understand the handling of their data.

After the lecture, students were asked to fill in a questionnaire that captured the constructs shown in Fig. 10.3 on a seven-point scale of rating, ranging from complete disagreement to complete agreement. The sample comprised 462 students, 65% of whom were male. The intended study programs were Business Administration (49%), Economics (22%), International Affairs (15%), Law (6%), and Law and Economics (8%). The average age of these first-semester students was 19.78 years (SD = 1.42 years).

The results of this study can be presented using an importance-performance map (Ringle & Sarstedt, 2016; Fig. 10.4).

The *x*-axis shows the strength of the association with the use intention (BI). For example, a value of 0.3 for adaptivity means that a perceived increase in the adaptivity of Lexi by one point on the seven-point rating scale yields an expected increase in the intention to use by 0.3 points. The *y*-axis of the importance-performance map shows the strength of the constructs on a percentage scale. For example, a value of 53 for adaptivity means that the perceived adaptivity is 53% of the theoretical maximum. The maximum would be achieved if students rate all adaptivity items at the highest possible value on the seven-point scale. Currently, the students in the sample rate the adaptivity of Lexi as medium. An importance-performance map enables users to identify constructs that are potentially promising for increasing BI. These would be constructs with a comparatively strong influence on BI, but at the same time a low performance or at least a performance that is well below 100%. In our study, all characteristics of the robot have a statistically significant positive influence on BI. However, the anxiety-related constructs (anxiety handling, anxiety backend) are not statistically significant and, in light of the statistical power of our study, do not have a practically relevant influence on BI.

Overall, the study showed that students tend not to have the intention to use Lexi as a learning aid—the performance of BI equals only 37% of the theoretical maximum. Following the idea of the importance-performance map, it would be promising to focus on the performance expectation, in other words, the perception of Lexi as a valuable learning aid. The current mediocre assessment is not surprising when research on high-quality learning arrangements (from the learner's perspective) is taken into account (Praetorius, Pauli, Reusser, Rakoczy, & Klieme, 2014). Lexi answered questions on a factual level which were not tailored to students' individual needs. Activities aimed at establishing a personal relationship between the lecturer and Lexi, such as using her name, were limited. Lexi also did not perform cognitive activation activities such as asking activating questions. Against this background, the students' assessment seems to be a realistic one. Substantially increasing PE seems to be difficult at the current technological state of the art. In addition, for personalized instruction it would be necessary to access student data, for example, from the learning management system, in order to retrieve student performance levels and then use this information in the interaction. If such a procedure is at all desirable, in light of privacy concerns, should be discussed.

SI could also be a factor in increasing the use intention of the robot. To this end, a communication strategy targeting the general public could be of value. For our part, we created a video introducing Lexi and its capabilities and distributed it via our social media channels and the website. The local press reported in a favorable way on the use of Lexi as a teaching assistant. Overall, it seems important to keep social influence in mind because it has a relatively strong impact on BI.

When looking at the characteristics of Lexi, adaptiveness stands out as the construct with the strongest influence on BI. As with human teachers, adaptivity seems to play an important role (Brühwiler & Blatchford, 2011). From the students' point of view, there is a clear need for improvement in adaptivity. This perception may also be realistic. However, it seems to be difficult to increase adaptiveness at the current technological state of the art, i.e., to consider students' prior knowledge, to provide learning material at an appropriate level of difficulty, to choose an appropriate learning pace, and to provide individualized feedback. To tackle this issue, it may be necessary to ensure that educational psychology and research on social robots and AI go hand in hand. One example could be the use of pupil dilation to measure cognitive load (van der Wel & van Steenbergen, 2018). Based on facial recognition (AI), the pupil diameter of students could be collected and the difficulty of the learning material adjusted accordingly to ensure appropriate cognitive load. Overall, it seems important to build on a strong conceptual basis for learning (Sweller, 2020). Afterward, it can be pointed out how AI can provide solutions that can be executed by social robots. However, it should be noted that well-trained human assistants would also probably not be able to achieve perfect adaptiveness.

Compared to adaptiveness, Lexi's other characteristics have a substantially lower influence on BI. Another remarkable finding is the low level of social presence (27%). The students do not have the impression that they are interacting with a real person. The findings regarding privacy concerns are also surprising. On the one hand, the students express strong concerns about privacy when interacting with the robot. On the other hand, these perceptions do not have an influence on their intention to use the robot. However, this may be explained by the "privacy paradox" (Acquisti, Brandimarte, & Loewenstein, 2015): people report serious concerns about privacy, yet voluntarily disclose private information and continue to use services, for example, social media, which they reportedly distrust. Lutz and Tamó-Larrieux (2020) found similar results for social robots.

In sum, at first glance the findings seem disappointing. Currently, there is considerable potential for increasing intention to use robots for learning purposes among university students of the social sciences. However, it was also possible to identify drivers that may be useful in increasing such user intention. Adaptiveness, in particular, seems to play an important role in the acceptance of social robots. As a limitation, it should be noted that the present study is based on correlative relationships. It cannot be ascertained whether causal effects are actually underlying these associations. Furthermore, it would be useful if students could work more intensively with robots in order to gain a better picture of the possibilities of social robots as learning aids.

10.5 Conclusions and Future Work

This chapter aims to provide an overview of the phenomenon of social robots in education. Evidence is available showing the value of physical presence. This is important because the higher cost involved in comparison to virtual agents has to be justified. The most commonly used type of robot in education is humanoid; a quasi-standard type is Nao from SoftBank Robotics. Social robots can interact with students using natural speech, motion, lights, and sounds. These characteristics contribute to the perceived personality of the robot. A further important social capability is empathy. Social robots can be described by their level of autonomy and intelligence. Autonomy is a continuum ranging from teleoperated to fully autonomous. A fully autonomous robot carries out sensing, planning, and acting without any intervention by the user. Since the desired level of autonomy is often hard to achieve at the current technological state of the art, researchers regularly apply the Wizard of Oz technique where the robot is teleoperated without the knowledge of the user. Due to the high complexity of educational settings, social robots have to be intelligent in order to achieve at least a moderate level of autonomy. Learning capabilities are regarded as a prerequisite for achieving intelligence.

In the classroom, various tasks have to be carried out which are codified in teaching standards. When identifying (sub-)tasks that can be undertaken by a robot, the criticality and complexity of the task need to be considered. Following the concept of hybrid intelligence, both the teacher and social robot may carry out the teaching tasks in collaboration and achieve a superior performance by utilizing the complementary strength of both parties.

The use of social robots has to meet ethical standards, namely, concerns about privacy, control, and responsibility. Moreover, AI-related issues such as the black-box problem and biased learning data have to be addressed. Besides ethical concerns, technology acceptance has to be considered; teachers, students, and parents, as important stakeholders, may be put into focus. Predictors for the intention to utilize social robots as learning assistants have been examined by presenting the case of academic writing; the perceived characteristic of the robot that best predicts the intention of use is the robot's adaptiveness.

Following the concept of hybrid intelligence, future research may not focus on robots in isolation but on how a teacher, in collaboration with the social robot, can best perform a specific task. The successful teacher may be the one who is competent in combining their own strength with that of the robot. When considering technology-related competencies, Seufert, Guggemos, and Sailer (2021), as well as Burkhard, Seufert, and Guggemos (2021), argue for the importance of such collaboration skills in an age of smart machines. For learning professionals in general, see Meier, Seufert, Guggemos, and Spirgi (2020). Future research could focus on the necessary knowledge, skills, and attitudes of pre- and in-service teachers (Seufert et al., 2021).

A promising avenue for further research could also be to explore how social robots can be integrated into a classroom ecosystem (Belpaeme & Tanaka, 2021; Seufert, Guggemos, & Moser, 2019). This would allow the generation of large-scale learning data that could be used to (further) train the robots. Moreover, ethical questions, for example, about privacy, can be addressed at the institutional level.

Finally, with social robots emerging as a new agent in the classroom, it might be promising to investigate how the orchestration of classroom activities could be effectively organized (Shahmoradi et al., 2019).

References

Abildgaard, J. R., & Scharfe, H. (2012). A geminoid as lecturer. In S. S. Ge, O. Khatib, J. J. Cabibihan, R. Simmons, & M. A. Williams (Eds.), *Social robotics—4th International Conference, ICSR 2012, Proceedings, Chengdu, China* (pp. 408–417). https://doi.org/10.1007/978-3-642-34103-8_41.

Acquisti, A., Brandimarte, L., & Loewenstein, G. (2015). Privacy and human behavior in the age of information. *Science, 347*(6221), 509–514. https://doi.org/10.1126/science.aaa1465

Admoni, H., & Scassellati, B. (2017). Social eye gaze in human-robot interaction: A review. *Journal of Human–Robot Interaction, 6*(1), 25. https://doi.org/10.5898/JHRI.6.1.Admoni

Ahmad, M. I., Mubin, O., & Orlando, J. (2017). Adaptive social robot for sustaining social engagement during long-term children–robot interaction. *International Journal of Human–Computer Interaction, 33*(12), 943–962. https://doi.org/10.1080/10447318.2017.1300750

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes, 50*(2), 179–211. https://doi.org/10.1016/0749-5978(91)90020-T

Alemi, M., Meghdari, A., & Ghazisaedy, M. (2015). The impact of social robotics on L2 learners' anxiety and attitude in English vocabulary acquisition. *International Journal of Social Robotics, 7*(4), 523–535. https://doi.org/10.1007/s12369-015-0286-y

Baraka, K., Alves-Oliveira, P., & Ribeiro, T. (2020). An extended framework for characterizing social robots. In C. Jost, B. Le Pévédic, T. Belpaeme, C. Bethel, D. Chrysostomou, N. Crook, et al. (Eds.), *Human-robot interaction* (pp. 21–64). Springer. https://doi.org/10.1007/978-3-030-42307-0_2

Bartneck, C., & Forlizzi, J. (2004). A design-centred framework for social human-robot interaction. In *Proceedings/RO-MAN 2004, 13th IEEE International Workshop on Robot and Human Interactive Communication: September 20–22, 2004, Kurashiki, Okayama, Japan* (pp. 591–594). IEEE Operations Center. https://doi.org/10.1109/ROMAN.2004.1374827.

Baxter, P., Ashurst, E., Read, R., Kennedy, J., & Belpaeme, T. (2017). Robot education peers in a situated primary school study: Personalisation promotes child learning. *PLoS One, 12*(5), e0178126. https://doi.org/10.1371/journal.pone.0178126

Beer, J. M., Fisk, A. D., & Rogers, W. A. (2014). Toward a framework for levels of robot autonomy in human-robot interaction. *Journal of Human–Robot Interaction, 3*(2), 74–99. https://doi.org/10.5898/JHRI.3.2.Beer

Belpaeme, T., Kennedy, J., Ramachandran, A., Scassellati, B., & Tanaka, F. (2018). Social robots for education: A review. *Science Robotics, 3*(21), 1–9. https://doi.org/10.1126/scirobotics.aat5954

Belpaeme, T., & Tanaka, F. (2021). Social robots as educators. In OECD (Ed.), *Digital education outlook 2021* (pp. 143–157). OECD. https://doi.org/10.1787/7fbfff45-en

Bransford, J., Brown, A. L., & Cocking, R. R. (2000). *How people learn: Brain, mind, experience, and school* (Expanded edition). National Academy Press.

Brščić, D., Kidokoro, H., Suehiro, Y., & Kanda, T. (2015). Escaping from children's abuse of social robots. In J. A. Adams, W. Smart, B. Mutlu, & L. Takayama (Eds.), *Proceedings of the 2015 ACM/IEEE International Conference on Human-Robot Interaction* (pp. 59–66). ACM. https://doi.org/10.1145/2696454.2696468

Brühwiler, C., & Blatchford, P. (2011). Effects of class size and adaptive teaching competency on classroom processes and academic outcome. *Learning and Instruction, 21*(1), 95–108. https://doi.org/10.1016/j.learninstruc.2009.11.004

Burkhard, M., Seufert, S., & Guggemos, J. (2021a). Paradigm shift in human-machine interaction: A new learning framework for required competencies in the age of artificial intelligence? In *13th International Conference on Computer Supported Education*. https://doi.org/10.5220/0010370300730083.

Burkhard, M., Seufert, S., & Guggemos, J. (2021b). Relative strengths of teachers and smart machines: Towards an augmented task sharing. In *13th International Conference on Computer Supported Education*. https://doi.org/10.5220/0010370300730083.

Byrne, B. M., Rossi, A., & Doolan, M. (2017). Humanoid robots attending lectures. In *Proceedings of the 9th International Conference on Education Technology and Computers, ICETC, Barcelona, Spain*. https://doi.org/10.1145/3175536.3175556

CCSSO. (2021). *InTASC model core teaching standards and learning progressions for teachers 1.0*. Council of Chief State School Officers (CCSSO). Retrieved from https://ccsso.org/resource-library/intasc-model-core-teaching-standards

Cheng, Y.-W., Sun, P.-C., & Chen, N.-S. (2018). The essential applications of educational robot: Requirement analysis from the perspectives of experts, researchers and instructors. *Computers & Education, 126*, 399–416. https://doi.org/10.1016/j.compedu.2018.07.020

Ching, Y.-H., Hsu, Y.-C., & Baldwin, S. (2018). Developing computational thinking with educational technologies for young learners. *TechTrends, 62*(6), 563–573. https://doi.org/10.1007/s11528-018-0292-7

Conti, D., Cirasa, C., Di Nuovo, S., & Di Nuovo, A. (2020). Robot, tell me a tale! *Interaction Studies, 21*(2), 220–242. https://doi.org/10.1075/is.18024.con

Cooney, M., & Leister, W. (2019). Using the Engagement Profile to design an engaging robotic teaching assistant for students. *Robotics, 8*(1), 1–26. https://doi.org/10.3390/robotics8010021

Dellermann, D., Ebel, P., Söllner, M., & Leimeister, J. M. (2019). Hybrid Intelligence. *Business & Information Systems Engineering, 61*(5), 637–643. https://doi.org/10.1007/s12599-019-00595-2

Eurobarometer 382. (2012). *Public attitudes towards robots*. European Commission.

Feuerriegel, S., Dolata, M., & Schwabe, G. (2020). Fair AI. *Business & Information Systems Engineering, 62*(4), 379–384. https://doi.org/10.1007/s12599-020-00650-3

Fong, T., Nourbakhsh, I., & Dautenhahn, K. (2003). A survey of socially interactive robots. *Robotics and Autonomous Systems, 42*(3–4), 143–166. https://doi.org/10.1016/S0921-8890(02)00372-X

Frey, C. B., & Osborne, M. A. (2017). The future of employment: How susceptible are jobs to computerisation? *Technological Forecasting and Social Change, 114*, 254–280. https://doi.org/10.1016/j.techfore.2016.08.019

Fridin, M., & Belokopytov, M. (2014). Acceptance of socially assistive humanoid robot by preschool and elementary school teachers. *Computers in Human Behavior, 33*, 23–31. https://doi.org/10.1016/j.chb.2013.12.016

Goel, A. K., & Polepeddi, L. (2016). *Jill Watson: A virtual teaching assistant for online education.* Retrieved from http://hdl.handle.net/1853/59104.

Graaf, M. M. A. d., & Allouch, S. B. (2013). Exploring influencing variables for the acceptance of social robots. *Robotics and Autonomous Systems, 61*(12), 1476–1486. https://doi.org/10.1016/j.robot.2013.07.007

Guggemos, J. (2021). On the predictors of computational thinking and its growth at the high-school level. *Computers & Education, 161*, 104060. https://doi.org/10.1016/j.compedu.2020.104060

Guggemos, J., & Seufert, S. (2021). Teaching with and teaching about technology—Evidence for professional development of in-service teachers. *Computers in Human Behavior, 115*, 106613. https://doi.org/10.1016/j.chb.2020.106613

Guggemos, J., Seufert, S., & Sonderegger, S. (2020). Humanoid robots in higher education: Evaluating the acceptance of Pepper in the context of an academic writing course using the UTAUT. *British Journal of Educational Technology, 51*(5), 1864–1883. https://doi.org/10.1111/bjet.13006

Gunning, D., Stefik, M., Choi, J., Miller, T., Stumpf, S., & Yang, G.-Z. (2019). XAI-Explainable artificial intelligence. *Science Robotics, 4*(37), eaay7120. https://doi.org/10.1126/scirobotics.aay7120

Handke, J. (2018). Digitale Hochschullehre – Vom einfachen Integrationsmodell zur Künstlichen Intelligenz. In U. Dittler & C. Kreidl (Eds.), *Hochschule der Zukunft* (pp. 249–263). Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-20403-7_15

Handke, J. (2020). *Humanoide Roboter: Showcase, partner und Werkzeug* (1st ed.). Tectum Wissenschaftsverlag.

Heerink, M., Kröse, B., Evers, V., & Wielinga, B. (2010). Assessing acceptance of assistive social agent technology by older adults: The Almere Model. *International Journal of Social Robotics, 2*(4), 361–375. https://doi.org/10.1007/s12369-010-0068-5

Henschel, A., Laban, G., & Cross, E. S. (2021). What makes a robot social? A review of social robots from science fiction to a home or hospital near you. *Current Robotics Reports, 2*(1), 9–19. https://doi.org/10.1007/s43154-020-00035-0

Hood, D., Lemaignan, S., & Dillenbourg, P. (2015). When children teach a robot to write. In J. A. Adams, W. Smart, B. Mutlu, & L. Takayama (Eds.), *Proceedings of the 2015 ACM/IEEE International Conference on Human–Robot Interaction* (pp. 83–90). ACM. https://doi.org/10.1145/2696454.2696479.

Huang, C.-M., & Mutlu, B. (2014). Learning-based modeling of multimodal behaviors for human-like robots. In G. Sagerer, M. Imai, T. Belpaeme, & A. Thomaz (Eds.), *HRI'14: Proceedings of the 2014 ACM/IEEE International Conference on Human-Robot Interaction* (pp. 57–64). ACM Press. https://doi.org/10.1145/2559636.2559668

Ifenthaler, D. (2015). Learning analytics. In J. M. Spector (Ed.), *The SAGE encyclopedia of educational technology* (pp. 447–451). SAGE. https://doi.org/10.4135/9781483346397.n187

Jamet, F., Masson, O., Jacquet, B., Stilgenbauer, J.-L., & Baratgin, J. (2018). Learning by teaching with humanoid robot: A new powerful experimental tool to improve children's learning ability. *Journal of Robotics, 2018*, 1–11. https://doi.org/10.1155/2018/4578762

Kärner, T., Warwas, J., Krannich, M., & Weichsler, N. (2021). How does information consistency influence prospective teachers' decisions about task difficulty assignments? A within-subject experiment to explain data-based decision-making in heterogeneous classes. *Learning and Instruction, 74*, 101440. https://doi.org/10.1016/j.learninstruc.2020.101440

Kennedy, J., Lemaignan, S., & Belpaeme, T. (2016). The cautious attitude of teachers towards social robots in schools. In *Robots 4 Learning Workshop at IEEE RO-MAN*.

Khanal, S. R., Barroso, J., Lopes, N., Sampaio, J., & Filipe, V. (2018). Performance analysis of Microsoft's and Google's Emotion Recognition API using pose-invariant faces. In *Proceedings of the 8th International Conference on Software Development and Technologies for Enhancing Accessibility and Fighting Info-exclusion* (pp. 172–178). ACM. https://doi.org/10.1145/3218585.3224223.

Leite, I., Castellano, G., Pereira, A., Martinho, C., & Paiva, A. (2014). Empathic robots for long-term interaction. *International Journal of Social Robotics, 6*(3), 329–341. https://doi.org/10.1007/s12369-014-0227-1

Li, J. (2015). The benefit of being physically present: A survey of experimental works comparing copresent robots, telepresent robots and virtual agents. *International Journal of Human-Computer Studies, 77*, 23–37. https://doi.org/10.1016/j.ijhcs.2015.01.001

Lutz, C., Schöttler, M., & Hoffmann, C. P. (2019). The privacy implications of social robots: Scoping review and expert interviews. *Mobile Media and Communication, 7*(3), 412–434. https://doi.org/10.1177/2050157919843961

Lutz, C., & Tamó-Larrieux, A. (2020). The robot privacy paradox: Understanding how privacy concerns shape intentions to use social robots. *Human–Machine Communication, 1*, 87–111. https://doi.org/10.30658/hmc.1.6

Mavridis, N. (2015). A review of verbal and non-verbal human–robot interactive communication. *Robotics and Autonomous Systems, 63*, 22–35. https://doi.org/10.1016/j.robot.2014.09.031

Meier, C., Seufert, S., Guggemos, J., & Spirgi, J. (2020). Learning organizations in the age of smart machines. In D. Ifenthaler, S. Hofhues, M. Egloffstein, & C. Helbig (Eds.), *Digital transformation of learning organizations* (pp. 77–94). Springer. https://doi.org/10.1007/978-3-030-55878-9_5

Mori, M., MacDorman, K., & Kageki, N. (2012). The Uncanny Valley. *IEEE Robotics and Automation Magazine, 19*(2), 98–100. https://doi.org/10.1109/MRA.2012.2192811

Mosavi, A., & Varkonyi-Koczy, A. R. (2016). Integration of machine learning and optimization for robot learning. In R. Jabłoński & R. Szewczyk (Eds.), *Advances in Intelligent Systems and Computing. Recent Global Research and Education: Proceedings of the 15th International*

Conference on Global Research and Education (pp. 349–355). Springer. https://doi.org/10.1007/978-3-319-46490-9_47.

Mou, Y., Shi, C., Shen, T., & Xu, K. (2020). A systematic review of the personality of robot: Mapping its conceptualization, operationalization, contextualization and effects. *International Journal of Human–Computer Interaction, 36*(6), 591–605. https://doi.org/10.1080/10447318.2019.1663008

Mousavinasab, E., Zarifsanaiey, N., Niakan, R., Kalhori, S., Rakhshan, M., Keikha, L., & Ghazi Saeedi, M. (2021). Intelligent tutoring systems: A systematic review of characteristics, applications, and evaluation methods. *Interactive Learning Environments, 29*(1), 142–163. https://doi.org/10.1080/10494820.2018.1558257

Mubin, O., Stevens, C. J., Shahid, S., Mahmud, A. A., & Dong, J.-J. (2013). A review of the applicability of robots in education. *Technology for Education and Learning, 1*(1), 1–7. https://doi.org/10.2316/Journal.209.2013.1.209-0015

Naneva, S., Sarda Gou, M., Webb, T. L., & Prescott, T. J. (2020). A systematic review of attitudes, anxiety, acceptance, and trust towards social robots. *International Journal of Social Robotics, 12*(6), 1179–1201. https://doi.org/10.1007/s12369-020-00659-4

Nasir, J., Norman, U., Bruno, B., & Dillenbourg, P. (2020). When positive perception of the robot has no effect on learning. In F. Ferland & D. Morocco (Eds.), *Robots with heart, mind, and soul: The 29th IEEE International Conference on Robot and Human Interactive Communication: 31 August-04 September 2020* (pp. 313–320). IEEE. https://doi.org/10.1109/RO-MAN47096.2020.9223343

Niculescu, A., van Dijk, B., Nijholt, A., Li, H., & See, S. L. (2013). Making social robots more attractive: The effects of voice pitch, humor and empathy. *International Journal of Social Robotics, 5*(2), 171–191. https://doi.org/10.1007/s12369-012-0171-x

Nishio, S., Ishiguro, H., & Hagit, N. (2007). Geminoid: Teleoperated android of an existing person. In A. C. de Pina Filho (Ed.), *Humanoid robots: New developments*. I-Tech Education and Publishing. https://doi.org/10.5772/4876

Pandey, A. K., & Gelin, R. (2018). A mass-produced sociable humanoid robot: Pepper: The first machine of its kind. *IEEE Robotics and Automation Magazine, 25*(3), 40–48. https://doi.org/10.1109/MRA.2018.2833157

Papadopoulos, I., Lazzarino, R., Miah, S., Weaver, T., Thomas, B., & Koulouglioti, C. (2020). A systematic review of socially assistive robots in pre-tertiary education. *Computers and Education, 155*, 103924. https://doi.org/10.1016/j.compedu.2020.103924

Pardo, A., & Siemens, G. (2014). Ethical and privacy principles for learning analytics. *British Journal of Educational Technology, 45*(3), 438–450. https://doi.org/10.1111/bjet.12152

Praetorius, A.-K., Klieme, E., Herbert, B., & Pinger, P. (2018). Generic dimensions of teaching quality: The German framework of three basic dimensions. *ZDM, 50*(3), 407–426. https://doi.org/10.1007/s11858-018-0918-4

Praetorius, A.-K., Pauli, C., Reusser, K., Rakoczy, K., & Klieme, E. (2014). One lesson is all you need? Stability of instructional quality across lessons. *Learning and Instruction, 31*, 2–12. https://doi.org/10.1016/j.learninstruc.2013.12.002

Reich-Stiebert, N., & Eyssel, F. (2015). Learning with educational companion robots? Toward attitudes on education robots, predictors of attitudes, and application potentials for education robots. *International Journal of Social Robotics, 7*(5), 875–888. https://doi.org/10.1007/s12369-015-0308-9

Reich-Stiebert, N., & Eyssel, F. (2016). Robots in the classroom: What teachers think about teaching and learning with education robots. In A. Agah, J.-J. Cabibihan, A. M. Howard, M. A. Salichs, & H. He (Eds.), *Social robotics* (pp. 671–680). Springer International Publishing.

Riek, L. (2012). Wizard of Oz studies in HRI: A systematic review and new reporting guidelines. *Journal of Human-Robot Interaction, 1*(1), 119–136. https://doi.org/10.5898/JHRI.1.1.Riek

Ringle, C. M., & Sarstedt, M. (2016). Gain more insight from your PLS-SEM results: The importance-performance map analysis. *Industrial Management and Data Systems, 116*(9), 1865–1886. https://doi.org/10.1108/IMDS-10-2015-0449

Salem, M., Kopp, S., Wachsmuth, I., Rohlfing, K., & Joublin, F. (2012). Generation and evaluation of communicative robot gesture. *International Journal of Social Robotics, 4*(2), 201–217. https://doi.org/10.1007/s12369-011-0124-9

Scherer, R., Siddiq, F., & Tondeur, J. (2019). The technology acceptance model (TAM): A meta-analytic structural equation modeling approach to explaining teachers' adoption of digital technology in education. *Computers and Education, 128*, 13–35. https://doi.org/10.1016/j.compedu.2018.09.009

Seeber, I., Bittner, E., Briggs, R. O., Vreede, T. d., Vreede, G.-J. d., Elkins, A., Maiera, R., Merza, A. B., Oeste-Reiße, S., Randrupf, N., Schwabeg, G., & Söllnereh, M. (2020). Machines as teammates: A research agenda on AI in team collaboration. *Information and Management, 57*(2), 103174. https://doi.org/10.1016/j.im.2019.103174

Serholt, S., Barendregt, W., Vasalou, A., Alves-Oliveira, P., Jones, A., Petisca, S., & Paiva, A. (2017). The case of classroom robots: Teachers' deliberations on the ethical tensions. *AI and Society, 32*(4), 613–631. https://doi.org/10.1007/s00146-016-0667-2

Seufert, Guggemos, & Moser. (2019). Digitale Transformation in Hochschulen: auf dem Weg zu offenen Ökosystemen. *Zeitschrift Für Hochschulentwicklung, 14*(2), 85–107. https://doi.org/10.3217/zfhe-14-02/05

Seufert, S., Guggemos, J., & Sailer, M. (2021). Technology-related knowledge, skills, and attitudes of pre- and in-service teachers: The current situation and emerging trends. *Computers in Human Behavior, 115*, 106552. https://doi.org/10.1016/j.chb.2020.106552

Seufert, S., & Spiroudis, E. (2017). Wissenschaftliche Schreibkompetenz in der Studieneingangsphase fördern: Erfahrungen mit einem Rückmelde und Bewertungsraster (Rubric). *Zeitschrift Schreiben, 15*, 1–9.

Shahmoradi, S., Olsen, J. K., Haklev, S., Johal, W., Norman, U., Nasir, J., & Dillenbourg, P. (2019). Orchestration of robotic activities in classrooms: Challenges and opportunities. In M. Scheffel, J. Broisin, V. Pammer-Schindler, A. Ioannou, & J. Schneider (Eds.), *Lecture notes in computer science. Transforming learning with meaningful technologies* (Vol. 11722, pp. 640–644). Springer International Publishing. https://doi.org/10.1007/978-3-030-29736-7_57

Sharkey, A. J. C. (2016). Should we welcome robot teachers? *Ethics and Information Technology, 18*(4), 283–297. https://doi.org/10.1007/s10676-016-9387-z

Smakman, M. H. J., Jansen, B., Leunen, J., & Konijn, E. A. (2020). Acceptable social robots in education: A value sensitive parent perspective. In *Proceedings of INTED2020 Conference* (pp. 7946–7953).

Smakman, M. H. J., Konijn, E. A., Vogt, P., & Pankowska, P. (2021). Attitudes towards social robots in education: Enthusiast, practical, troubled, sceptic, and mindfully positive. *Robotics, 10*(1), 24. https://doi.org/10.3390/robotics10010024

Sohn, K., & Kwon, O. (2020). Technology acceptance theories and factors influencing artificial intelligence-based intelligent products. *Telematics and Informatics, 47*, 101324. https://doi.org/10.1016/j.tele.2019.101324

Spolaôr, N., & Benitti, F. B. (2017). Robotics applications grounded in learning theories on tertiary education: A systematic review. *Computers and Education, 112*, 97–107. https://doi.org/10.1016/j.compedu.2017.05.001

Suchman, M. C. (1995). Managing legitimacy: Strategic and institutional approaches. *Academy of Management Review, 20*(3), 571–610. https://doi.org/10.5465/amr.1995.9508080331

Sweller, J. (2020). Cognitive load theory and educational technology. *Educational Technology Research and Development, 68*(1), 1–16. https://doi.org/10.1007/s11423-019-09701-3

Tanaka, F., Isshiki, K., Takahashi, F., Uekusa, M., Sei, R., & Hayashi, K. (2015). Pepper learns together with children: Development of an educational application. In T. Asfour (Ed.), *15th International Conference on Humanoid Robots (Humanoids), IEEE-RAS: November 3–5, 2015, Seoul, Korea* (pp. 270–275). IEEE. https://doi.org/10.1109/HUMANOIDS.2015.7363546

Vallor, S., & Bekey, G. A. (2017). Artificial intelligence and the ethics of self-learning robots. In P. Lin, R. Jenkins, & K. Abney (Eds.), *Robot Ethics 2.0: From autonomous cars to artificial intelligence* (pp. 338–353). Oxford University Press.

Van den Berghe, R., Verhagen, J., Oudgenoeg-Paz, O., van der Ven, S., & Leseman, P. (2019). Social robots for language learning: A review. *Review of Educational Research, 89*(2), 259–295. https://doi.org/10.3102/0034654318821286

Van der Wel, P., & van Steenbergen, H. (2018). Pupil dilation as an index of effort in cognitive control tasks: A review. *Psychonomic Bulletin & Review, 25*(6), 2005–2015. https://doi.org/10.3758/s13423-018-1432-y

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly, 27*(3), 425–478. https://doi.org/10.2307/30036540

Vogt, P., van den Berghe, R., Haas, M. de, Hoffman, L., Kanero, J., Mamus, E., Montanier, J.-M., Oranc, C., Oudgenoeg-Paz, O., Garcia, D. H., Papadopoulos, F., Schodde, T., Verhagen, J., Wallbridge, C. D., Willemsen, B., de Wit, J., Belpaeme, T., Göksun, T., Kopp, S., Krahmer, E., Küntay, A. C., Leseman, P., & Pandey, A. K. (2019). , Second language tutoring using social robots: A large-scale study. In *Proceedings of the 14th ACM/IEEE International Conference on Human-Robot Interaction* (pp. 497–505). IEEE Press. https://doi.org/10.1109/HRI.2019.8673077.

Wang, P. (2019). On defining artificial intelligence. *Journal of Artificial General Intelligence, 10*(2), 1–37. https://doi.org/10.2478/jagi-2019-0002

Woo, H., LeTendre, G. K., Pham-Shouse, T., & Xiong, Y. (2021). The use of social robots in classrooms: A review of field-based studies. *Educational Research Review, 33*, 100388. https://doi.org/10.1016/j.edurev.2021.100388

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education—Where are the educators? *International Journal of Educational Technology in Higher Education, 16*(1). https://doi.org/10.1186/s41239-019-0171-0

Zhong, B., & Xia, L. (2018). A systematic review on exploring the potential of educational robotics in mathematics education. *International Journal of Science and Mathematics Education, 12*(5), 491. https://doi.org/10.1007/s10763-018-09939-y

Chapter 11
Undergraduate Mathematics Students Engaging in Problem-Solving Through Computational Thinking and Programming: A Case Study

Said Hadjerrouit and Nils-Kristian Hansen

11.1 Introduction

Students in mathematics study programs are expected to acquire basic algorithmic and computational thinking (CT) skills, in addition to learning emerging programming languages (Wing, 2014). These are considered important competencies for future work in society and should be acquired by all university mathematics students to improve their mathematical problem-solving skills by benefitting from CT and the power of programming languages (Shute, Sun, & Asbell-Clarke, 2017). It is assumed that with easier access to digital technology in higher education, the integration of programming activities and application of CT skills into teaching and learning could be easier than in previous years. Indeed, until the relatively recent re-emergence of interest for integrating CT and programming in school mathematics, this topic received relatively little attention in the research community. Still, despite the renewed interest, there is little research on linking CT and programming with mathematics at the university level (Buteau, Muller, Mgombelo, Sacristán, & Dreise, 2020).

This study is an attempt to *contribute* toward the advancement of the potential value of CT at the undergraduate level when students engage in mathematical problem-solving through CT and programming, while uncovering both opportunities and challenges encountered by the students when trying to connect mathematics with CT and programming.

This chapter is structured as follows. Firstly, the theoretical framework is outlined. Secondly, the chapter describes the context of the study, research question, methods, and the mathematical task to be solved. Then, the results are reported and

S. Hadjerrouit (✉) · N.-K. Hansen
University of Agder, Kristiansand, Norway
e-mail: Said.Hadjerrouit@uia.no

D. Ifenthaler et al. (eds.), *Orchestration of Learning Environments in the Digital World*, Cognition and Exploratory Learning in the Digital Age,
https://doi.org/10.1007/978-3-030-90944-4_11

analyzed. This is followed by a discussion and pedagogical implications of the results. Finally, the limitations of the study are highlighted, and recommendations for future work conclude the chapter.

11.2 Theoretical Framework

The application of computational thinking (CT) and programming for mathematical problem-solving has had a relatively long history in mathematics education (Papert & Harel, 1991; Shodiev, 2015). Today, CT and programming languages are becoming a key learning goal of mathematics courses from primary to university level. Drawing on the research literature, the theoretical framework and associated terms and notions underlying this study are outlined in the following sections.

11.2.1 Computational Thinking (CT): A Review of the Research Literature

While a large volume of research studies on CT and programming in mathematics at the school level exists, there is little research at the university level on linking CT and programming with mathematics. Some important research issues and challenges have been identified.

Broley, Caron, and Saint-Aubin (2018) reported that teachers have different understanding and interpretations of CT and programming. They do not associate it automatically with a systematic and well-organized way of mathematical problem-solving with best possible outcome. Moreover, there may be numerous justifications for exposing students to CT and programming depending on the educational situation.

Similarly, Li et al. (2020a, 2020b) argued for the complexity of CT, which is not synonym with "computing" or "computer" in a restricted sense. Rather, CT is a model of thinking with a multifaceted theoretical perspective. CT is important not only in computer science and mathematics but also in other disciplines of STEM.

In more general terms, Malyn-Smith and Angeli (2020) distinguished between two definitions of CT, one focusing on defining CT by disaggregating its elements and another on exploring the integration of CT into disciplinary learning through its practices. While there are similarities between these two definitions, there are several differences. Accordingly, the challenge includes the evolution of a common definition of CT, and a shared understanding of CT, comprehensively integrating CT into curricula.

In terms of theoretical frameworks, Buteau, Muller, Marshall, Sacristán, and Mgombelo (2016) discussed CT and programming from a broad perspective based on Wing (2008) and other researchers, and from the perspective of mathematicians' research practices, partly informed by the constructionist paradigm (Papert & Harel, 1991), before they developed their theoretical framework on their view of learning mathematics by engaging in CT, drawing mainly on the instrumental approach.

Similarly, Buteau et al. (2018) used the constructionist approach for the classroom implementation of programming and the instrumental approach to address a student's appropriation of CT and programming as a tool for the exploration of a mathematics concept, theorem, conjecture, or a real-world situation. See also Buteau et al. (2020) for the use of instrumental genesis (an important element of the instrumental approach) stages of programming for mathematical work.

Gueudet et al. (2020) also used the instrumental approach to highlight the links and connections between mathematical and programming competencies, which are quite complex and increasingly important at the university level. The study recommends deepening the knowledge about these complex links and their evolution, including CT even though this notion is not mentioned in the article.

In line with Gueudet et al. (2020), an overview of research done by DeJarnette (2019) concluded that the question that is still underdeveloped in existing literature on CT at the university level is how students develop skills when interacting within an environment that merges and connects mathematical thinking with CT and programming.

Summarizing, the research literature reveals important issues and challenges to be addressed for the advancement of mathematical problem-solving through CT and programming and the importance of integrating these skills into mathematics courses and work of today's mathematician (Broley et al., 2018). This study aims to address some of these issues and challenges, in particular the complex links between mathematics, CT, and programming.

11.2.2 Computational Thinking (CT) and Mathematical Thinking

Several definitions of the term CT exist in the research literature. Wing (2014) described CT as "the thought processes involved in formulating a problem and expressing its solution(s) in such a way that a computer—human or machine—can effectively carry out." Misfeldt and Ejsing-Duun (2015) described CT in similar words. It is the ability to work with algorithms understood as systematic and structured descriptions of problem-solving and construction strategies. Filho and Mercat (2018) defined algorithmic thinking as the process of solving a problem step by step in an effective, nonambiguous, and organized way that can be translated into instructions to solve problems of the same type by an individual or a computer.

Wing (2008, 2014) indicated that the main commonality between CT and mathematical thinking is problem-solving and a structured step-by-step construction process. Likewise, mathematical thinking consists of problem-solving processes, beliefs about mathematics, and justification for solutions (Shute et al., 2017, p. 145).

Accordingly, it involves the application of mathematical skills to solve mathematical problems, e.g., equations and functions (Shute et al., 2017, p. 145). The authors concluded that mathematical thinking and CT have a lot of communalities: problem-solving, modeling, data analysis and interpretation, as well as statistics and probability (Shute et al., 2017, p. 145). Furthermore, CT has communalities with engineering thinking in terms of design and evaluation of processes (Pérez-Marín, Hijón-Neira, Bacelo, & Pizarro, 2020), which is similar to algorithmic thinking and design of programming code. Moreover, CT and programming constructs such as variables and flow statements (if-then-else, for, while, repeat, etc.) are closely connected to arithmetical and mathematical thinking (Lie, Hauge, & Meaney, 2017).

This close connection between mathematical thinking and CT might provide opportunities for mathematical problem-solving. In contrast, programming skills alone without the mediation of CT are important but may not be sufficient to improve mathematical problem-solving. Clearly, CT is not the same as programming, but being able to program and test program codes is a result of being able to think computationally (Li et al., 2020a, 2020b; Shute et al., 2017; Wing, 2008). Thus, CT skills are critical for building efficient algorithms for mathematical problem-solving rather than trial and error and getting the program to run (Topalli & Cagiltay, 2018). In other words, CT requires students to be engaged in a continuously changing problem-solving process until an appropriate solution is found by designing effective algorithms that can be translated into computer programs.

11.2.3 A Three-Step Iterative Approach to Mathematical Problem-Solving Through CT and Programming

Drawing on the research literature (Kotsopoulos, Floyd, Nelson, Makosz, & Senger, 2019; Lee & Malyn-Smith, 2020; Romero, Lepage, & Lille, 2017; Santos, Tedesco, Borba, & Brito, 2020; Weintrop et al., 2016), this chapter proposes a three-step iterative approach to connect mathematical problem-solving with CT and programming (Fig. 11.1).

Firstly, students should have a good mathematical background to benefit from CT and programming languages. More specifically, they should be able to benefit from their knowledge to make sense of the task and have a mathematical understanding of it before formulating an algorithm and starting programming. Secondly, CT should enable students to analyze and decompose the mathematical task into smaller sub-tasks, analyze them in a different way than one would otherwise do in educational settings, and design an algorithm on how to solve it step by step before programming it. Engaging students in mathematical problem-solving through CT may enable a better understanding of mathematics beyond textbook mathematics and paper-pencil techniques. Thirdly, students should be able to translate the mathematical solution with the associated algorithm to the constructs of the programming language. This presupposes that the language is usable for novice students.

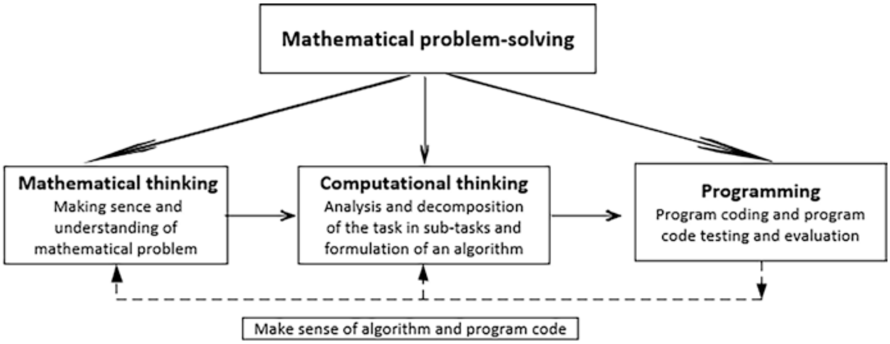


Fig. 11.1 Three-step approach to connect mathematical problem-solving with CT and programming

This is not a linear approach beginning with a mathematical problem and ending with programming and code testing. The approach may include numerous feedbacks to previous steps to make sense of the algorithm and program output. It is also particularly important to consider interactions between mathematical thinking and CT. Using CT and programming in mathematics courses should provide opportunities to help students do mathematics and gain new knowledge that is otherwise difficult to acquire without experimenting with the program and thinking algorithmically. However, this might be difficult to achieve for novice students unless the mathematical tasks are well designed, mathematically sound, and faithful to the underlying mathematical properties.

11.2.4 Programming and Usability Issues

When referring to the notion of usability or usable technologies in mathematics education, the research literature focuses on educational software such as GeoGebra, CAS, and SimReal (Artigue et al., 2009; Bokhove & Drijvers, 2010; Hadjerrouit & Gautestad, 2019). However, programming languages differ from educational software and how they are used to implement mathematical problems where one, for example, can graph a function or compute an integral simply by entering the function and pressing a button in GeoGebra or similar. Hence, evaluating the usability of programming languages might not be as straightforward as it may seem. Still, three usability criteria can be applied to programming languages with slight modification.

The first criterion is the degree to which the user interface of the programming language is easy to use and understand. Secondly, a usable programming language should allow a quick familiarization with it in terms of learning the language constructs, such as variables, if-then-else, for and while loop, or repeat. The third criterion is the quality of feedback provided by the program in terms of semantic and syntax error messages and whether these are useful to foster a successful implementation of the algorithmic solution through testing, correcting and improving the program, and making sense of the program output in terms of the mathematical solution to the problem.

These usability issues are connected to each other. Interface usability and familiarization with the programming language constructs are a prerequisite for successful implementation of the program and making sense of the solution. Indeed, technical programming constraints may result in demotivation and frustration from using the programming language. This can happen if there are technical usability issues or the program is not well designed. It can then be difficult to detect runtime or syntax errors, even if the programming language being used—MATLAB—comes up with hints. As a result, students may not work on their own if the feedbacks from the program are not comprehensive.

11.2.5 CT, Programming, and Pedagogical Considerations

A purely technical approach to CT and programming will not succeed unless students' engagement with mathematical problem-solving is placed in a pedagogical context. A pedagogy-based approach to CT and programming should enable a good degree of autonomy so that the students can work on their own and have a sense of control over their mathematical learning. Clearly, students should be able to acquire knowledge without being completely dependent on the teacher. Moreover, CT and programming languages should be a motivational factor for learning mathematics. In other words, CT should support students' engagement with problem-solving by means of intrinsically motivating tasks that are tied to the students' mathematical activities. Finally, interacting with a programming language when engaging in mathematical problem-solving should be mediated by CT through a structured construction of effective algorithms.

11.3 The Study

11.3.1 Context of the Study and Research Question

This work is a single case study conducted in the context of a first-year undergraduate course on programming with applications in mathematics. The participants were two students volunteering from a class of 8, enrolled in the course in 2019. The students had varied knowledge background in mathematics, but no experience with the programming language MATLAB. The course introduced the basic constructs of MATLAB, e.g., single variables, arrays, control flow statements, and functions. The course also discussed the major steps in systems development, i.e., analysis, design, implementation, and testing. Ultimately programming was used for numerical analysis. No explicit training on CT was provided, though. Student programming exercises made up a major part of the course. However, they had no focus on CT.

The research question addressed in this chapter is*: How do students engage in mathematical problem-solving through CT and programming activities?*

11.3.2 Mathematical Task

The mathematical task presented to the students in this research study was as follows:"The length of a curve may be approximated using Pythagoras' theorem by positioning a triangle adjacent to the curve (Fig. 11.2, left, below). The length of the green line between A and B may then be approximated as $\sqrt{x^2 + y^2}$. The task is to write a MATLAB function approximating the curve length of $f(x) = 2^x$ between two given x-values (Fig. 11.2, right, below).

A function skeleton is as follows:

- Function length = length_estimate(x1, x2),
- Length =?;

Determine the formula, based on x1 and x2, to replace the question mark. To calculate the square root, you may use the MATLAB function sqrt. The command sqrt(x) will produce $\sqrt{x}$."

Approximating curve length in this fashion is a mathematically sound first approach. If extended to a sum of an ever-increasing number of ever smaller triangles, it will converge to the actual curve length.

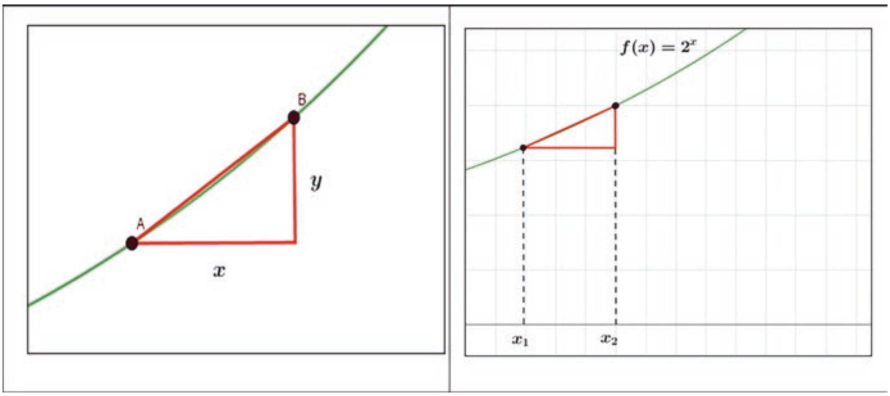


Fig. 11.2 The mathematical task

11.3.3 The Programming Language MATLAB

MATLAB is a mathematics software with a built-in C-based scripting language. A screenshot of the script editor containing a suggested solution to the task is shown in Fig. 11.3.

This is very similar to C, except for that in MATLAB syntax the function return variable, e.g., "length" in Fig. 11.3, is located in the function header instead of in a separate return statement.

11.3.4 Data Collection and Analysis Method

The two participating students were presented with a mathematical task to solve, while responding to questions in a dialogue with the teacher on the mathematical solving process. The main data collection method used is a task-based semi-structured interview with the students. Open-ended questions were also used to gain a deeper understanding of some important issues.

In terms of data analysis of the results, the three-step approach presented in the theoretical framework (Sect. 2.3) served as a reference for analyzing the students' problem-solving process, that is:

1. Understand the mathematical task.
2. Analyze and decompose the mathematical task, and then design an algorithm on how to perform the solution step by step before programming it.
3. Finally, translate the algorithmic solution to the MATLAB programming language code, with eventual feedbacks to previous steps 1 and 2.

More specifically, the students were expected to solve the mathematical task presented to them in Sect. 3.2. in three steps as follows:

- Understand the task, that is, using Pythagoras' theorem to calculate the hypotenuse.
- Formulate an algorithm, that is, find the lengths of the triangle hypotenuse using the function $f(x) = 2^x$, relating it to $x1$ and $x2$.
- Translate the algorithm into MATLAB code, corresponding to the sample shown in Fig. 11.3.

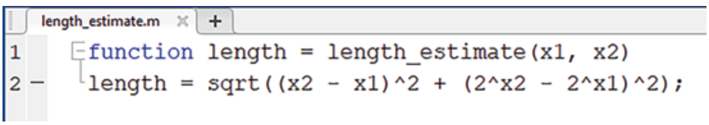


Fig. 11.3 MATLAB script editor with task solution

The analysis of the results seeks indications of students' problem-solving through CT by means of MATLAB according to this three-step approach. This is not the same as analyzing and coding in the sense of grounded theory without theoretical background. Rather, it is an analytical tool that tries to address the research question about how students engage in mathematical problem-solving through CT and programming activities drawing on students' interviews when solving the mathematical task. The interviews were transcribed and analyzed according to an inductive strategy based on the interplay between the three-step approach and the empirical data collected by means of semi-structured interviews (Patton, 2002).

11.4 Results

The results describe how the participating students engage in mathematical problem-solving through CT and MATLAB. The students were given the task described in Sect. 3.2. The abbreviations S1 and S2 are used for the students, and T for the teacher.

11.4.1 Student 1 (S1)

At the time of the interview, the student almost had completed the first semester of a bachelor program in mathematics. The interview took place in a classroom with the student, the teacher, and an observer present. Audio from the session was recorded and the dialogue later transcribed. The student brought own computer and used it to program in MATLAB. A piece of paper and a pen was put in front of the student with instructions to use if desired. The task was presented to the student on a sheet of paper as the interview started.

It took some time before the student made sense of the mathematical task. The teacher then asked the student to develop a skeleton of the solution. The student did, and started thinking about the length, but suggested an incorrect solution, based only on the values of $x1$ and $x2$:

S1: I first sat and thought function length, estimating length, x1, x2 and then I thought length is equal to the square root of x1 to the power of two plus x2 to the power of two.

At this stage, the student showed a basic mathematical and algorithmic understanding of the task, by referring to the mechanisms of Pythagoras' theorem and linking it to the variables available in the MATLAB function. However, the student failed to see the need for including the function $f(x) = x^2$ in the calculations.

After some calculation trials, the student noticed that the attempted solution was wrong. Then, the teacher encouraged the student to think computationally:

T: *But before you start using MATLAB, are you going to make an algorithm (…), for problem-solving before you start using MATLAB?*
S1: *I just have to sit and think about it.*

But still, the student continued guessing and calculating without thinking computationally. Neither did the student attempt to make use of pen and paper as an aid in structuring the thought processes, for instance, by drawing a sketch. After a while, the student's mental model reversed, and the student attempted to calculate the known x and y coordinates from the unknown hypotenuse. At this point the teacher intervened:

T: Now you say you know the hypotenuse and calculate y. But the hypotenuse is the unknown parameter here, the one you are supposed to calculate.
S1: Yes.
T: So now you have turned the problem around (…). It is just that that thought was a little backwards, maybe.
S1: Yes, it is quite possible.

Following this short dialogue, the student started using MATLAB without developing an algorithmic solution or a clear strategy for solving the problem. The teacher then engaged in a discussion guiding the student step by step toward an algorithmic solution and a MATLAB function, linking the program code to the mathematical task by continuously referring to Fig. 11.2. Eventually the program was tested with $x1 = 0$ and $x2 = 1$ and output 14,142. The teacher then wanted the student to reflect on the fact that the number probably was the square root of 2, but got no response. The student thus failed to make mathematical sense of the output, realizing that the test coordinates in question would result in $x = 1$ and $y = 1$ as in Fig. 11.2 and thus a hypotenuse of $\sqrt{1+1} = \sqrt{2}$.

Referring to Fig. 11.1, an iterative step-by-step approach from mathematical task to tested program code now had been completed. The teacher then wanted the student to reflect on the process just undertaken:

T: (…). Do you have anything to say about (…) like that afterwards?
S1: No, I am, I was a little bit in doubt about how to (…) First, it was the task you asked about (…) and then it was (…) and then I thought (…) f(x) is the function in x2 would be that point minus the function of that point (…) that it would be the length. But that is where I was wrong, I felt (…) Because you meant it to be here (…) and I understand that now.
T: Yes, that is the point, (…). That is why you have to use Pythagoras to find (…). Did you think (…) There was a hint here, wasn't there? Square root?
S1: Yes, yes, yes, the square root (…). I knew it was probably wrong, but I just did not quite understand what it was.
T: Well, because there was a clue there that you were unable to use, wasn't there. Then you realize that there is something (…)

This excerpt shows there is little indication that the student was following a clear problem-solving strategy, which confirms that making sense of the problem and having a mathematical understanding of the task is of crucial importance before formulating an algorithm and starting programming, which is not the case in this excerpt. A few minutes later, the teacher asked the student if there is a tendency to favor pen and paper to solve the task algorithmically before starting using MATLAB since developing an algorithm does not automatically require using the computer:

S1: If I have it in my head, sometimes I start with MATLAB, and then I write some sort of sketch before going through it carefully. If I am not sure, I will start with paper.
T: Maybe the task was not quite clear?
S1: Yes, so far, but I had probably forgotten some of the principles there.
T: Principles related to MATLAB or to the mathematical assignment?
S1: To the mathematical problem.

Again, the interview shows that the problem-solving process requires a good understanding of the mathematical task and computational thinking skills before programming it. Here again, the teacher reminded the student about the importance of algorithmic thinking before translating the task into MATLAB code:

T: Now, the point of the assignment is that you should be able to translate the mathematical task into code in MATLAB. That is really the point here.
S1: Yes, I felt that when I understood the mathematical thing, I had no trouble putting it into MATLAB. It was simply that I had (…) forgotten a bit the length thing there. That f of that minus f of that is delta y, then.

11.4.2 Student 2 (S2)

This student was an in-service mathematics teacher taking the course as a self-study. The student lived far from campus, and the interview therefore took place in the video conferencing system Zoom (https://zoom.us/) with only the teacher and the student present. The session was recorded, and the dialogue later transcribed. The task was sent to the student by e-mail as the interview started.

After a short dialogue with the student on the menu structure of MATLAB, which the student found unappealing, the student started studying the task. The teacher motivated by pointing out that this kind of exercise lately had become part of the exams in the programming course. Then, the teacher invited the student to reflect on the solving process:

T: I just have to ask you … How will you proceed to attack such a task? Will you go straight for the keyboard, and then …? You see, you have got a small skeleton here; will you go straight for the keyboard and start programming it, or will you …
S2: No, first I really try to think it through carefully, what the task is really about, here. And then I tend to make a draft as a first thing. (…) That I write down what the assignment actually asks for, and then that I make myself, in a way, a kind of design, then.

T: A design, yes.

S2: And what is it that the program is supposed to be able to calculate here, then? And then, when it is plainly clear to me, I would have gone into more depth on … on … what is somehow … how I am supposed to describe the code here. And here, in this case, there is a function, which is an estimate of length. And then I had to go in and have a look at the length. And then put it in a formula … Oh, I do not know if you want me to do the task itself, kind of.

T: (…) What we are a looking for, is the thought process. You say you create a draft, do you write code then, or draw boxes, or?

S2: Sometimes I have drawn one of these … in a way a line, then. And then I have drawn a box. Or, if it involves if-statements, then I have drawn something a bit like this. (Shows sketch like Fig.11.4, left, below.) If you look here. And then I have sort of a condition inside here, and then I have, if there is an if at the conditions, then I put this there (Shows sketch like Fig.11.4, right, below, pointing to the new line.)

T: That is what we call a flow chart.

S2: Flow chart, simply.

T: Yes, that is probably something you have learned once, maybe.

S2: Yes, I learned it in this study.

This excerpt shows that the student was aware of the importance of designing an algorithmic solution before starting to program the task, in stark contrast to student 1 who struggled to understand the mathematical task and formulating an algorithm. Words like "design," "draft," "box," and "function" point to algorithmic and computational thinking. Associating "if-conditions" with a flowchart also demonstrates the ability to create an abstract model from programming language construct.

Furthermore, the student makes it clear that employing CT before starting to program is a practice that has developed over time, the process being motivated by personal experience:

T: Yes, okay, yes. Very good, for it has turned out that people are very slow to adopt … They often go straight on to the programming. And that works fine when

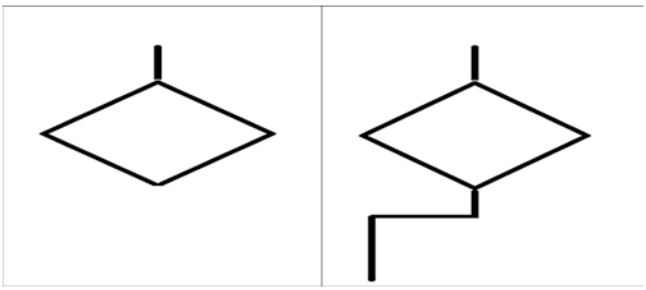


Fig. 11.4 Student's sketches

the task is easy, but when it gets a little more complex, then you need to split things up a bit.

S2: Yes, I had not thought about it until I, in a way, read it in this study, but after that I have started to … then I spend a bit of time in the process before I start typing the formula. If I have already started typing the formula, and then want to change it afterwards, then I have discovered that I do a lot of mistakes. So now I have tried to spend plenty of time on the preparatory work, so that I am one hundred per cent sure of what it is my function or program is supposed to do. And now I know it is a function, I am supposed to write, and what is it … what is it I am actually supposed to calculate here?

In this excerpt, the student demonstrates the ability to reflect upon own practice and followingly to abstract and articulate the essence of the process, mentioning the CT concept of dividing a task into smaller subtasks:

T: (…)

S2: It is the kind of like you analyze this pretty carefully, and then you divide it, and then … in a way … I do not know, I work kind of structured, then, with one piece at a time.

T: Yes, but that is really good. It is a concept one has in programming, you make a design, then you divide it into pieces, where you can look at one piece at a time.

Then, the student continued explaining the way of thinking algorithmically and computationally, including the programming process and the testing of the program:

S2: And then after I have made … made one, then … if it … now this task was a little bit different, then … but … but if I have made a program … after I have written the entire program code … If it typically is with for-loops and everything, then I always run through the program in my head. And write down that, okay, now, n is equal to 1, What happens to that and that value. In that way I always get to put it to the test, and then I detect possible errors.

Referring to the step-by-step process indicated in Fig. 11.1, the student here switches back and forth between CT and programming, by alternating between a mental analysis of the mechanisms in the algorithm and testing the actual program code.

11.5 Discussion

This study aimed to address this research question: *How do students engage in mathematical problem-solving through computational thinking and programming activities?* The results show that the introduction of CT and programming skills into the undergraduate level presents many challenges and opportunities for students committed to improving their understanding of mathematics. The analysis of the results shows that the participating students had two opposite experiences. They engaged differently in the mathematical problem-solving process. While the first student (S1) felt challenged by the mathematical task and the way to handle it through CT and programming, the second student (S2) made good use of CT, algorithms, and MATLAB to solve the task. The main finding of this study is that explicitly linking mathematical thinking with CT and programming is the key factor to ensure a successful implementation of the three-step iterative approach to mathematical problem-solving and enhance students' understanding of mathematical concepts.

11.5.1 Summary of the Results

For student 1 (S1), the first challenge was the lack of ability to create a mental mathematical model of the problem, hindering the student in making sense of the task, and then developing a problem-solving strategy translatable into an algorithm. This illustrates the approach presented in Sect. 2.3, requiring students to have a good mathematical understanding of the task as a basis for further development and use of computational and algorithmic thinking skills. The second challenge was related to the implementation of the algorithm using the constructs of the programming language. The student was unable to relate it to the mathematical task using an algorithm. Neither was the student able to make mathematical sense of a result output by the MATLAB program. Finally, once the teacher had guided the student step by step through the task, the student was unable to recapitulate the process. Clearly the problem-solving process in three steps outlined in the theoretical framing was challenging for the student, even after completing a programming course.

Student 2 (S2) was on the other hand very prepared to make use of CT. Student 2 had sufficient mathematical modeling capacity to make sense of the task and then to design a structured problem-solving solution translatable into an algorithm. Moreover, in contrast to student 1, student 2 demonstrated good ability to reflect, abstract, and structure and was able to use these abilities to analyze the mathematical task and to formulate an algorithm. This student was an in-service mathematics teacher and thus supposedly was trained in mathematical modeling and problem-solving prior to taking the programming course. Referring to the approach presented in Sect. 2.3, the mathematical abilities may have propagated through the steps in the model, facilitating the use of CT and programming. This is in contrast to student 1, where struggling with making sense of the mathematical task was a hindrance in going through the other steps in the approach.

11.5.2 Pedagogical Implications

Several pedagogical implications for teaching and learning CT for mathematical problem-solving can be drawn from the results. Firstly, the opposite students' experiences show that the minimum requirement for applying CT to mathematical problem-solving is a good combination of background knowledge in mathematics, algorithmic thinking skills, familiarity with the programming language constructs in question, program testing and validation, and making sense of the program output. Clearly, if students have the basic mathematical understanding required, they would be able to easily analyze mathematical tasks and break them down into small sub-tasks, designing an algorithm before moving on to programming as student 2 demonstrated in this study. These requirements and ways of problem-solving should be supported and implemented to become integral parts of university mathematics courses.

Secondly, to ensure a successful integration of CT into mathematics courses, the pedagogical setting around these courses should be well designed in terms of varied and intrinsically motivating tasks that reflect students' knowledge level. A good integration of CT into mathematics courses should also promote student autonomy and ownership. However, autonomy cannot be fully expected for novice students without good knowledge background in mathematics and some familiarities with CT and programming language constructs. Clearly, CT is a difficult matter for novice students, because it is more a way of thinking than a computational skill to acquire.

Thirdly, as this study shows and following the second implication, the role of the teacher as a facilitator of learning is still crucial to assist novice students in mathematical problem-solving by means of CT and programming. The role of a knowledgeable teacher is to foster CT and connections between mathematical thinking, programming, and algorithmic thinking.

Finally, following the second and third implication, a pedagogical setting that promotes a learning environment capable of motivating students should be created around mathematical problem-solving, CT, and programming practices. In other words, to alleviate the difficulty of thinking computationally, the three-step approach presented in this study needs to be combined with a pedagogically sound model, which emphasizes two elements: The teacher as facilitator of learning, and peer collaboration to allow the more competent students help those facing difficulties in accomplishing the computational tasks (Kuo & Hsu, 2020). Research on peer collaboration may contribute to explore collaborative learning occurring in the process of learning CT and programming.

11.5.3 Limitations of the Study

While the results from the opposite students' experiences and insights of the present study contribute to offer a better understanding on the interactions between CT, mathematical thinking, and programming, the study is limited to two students and can therefore not be generalized to all students enrolled in the course. The study is also limited to a very specific mathematical task (Pythagoras' theorem) involving undergraduate students with no prior experience in CT and programming. A final limitation that may have influenced the study is the method being used to gather data, that is, the semi-structured interview with the students once they have finished the course, one performed online.

11.6 Conclusions and Future Research

Although the participating students may not be representative for the average student enrolled in the course and the task was specific, two preliminary conclusions can be drawn from the study. Firstly, the connections between mathematics, CT, and programming languages are quite complex in line with the research literature (Li et al., 2020a, 2020b) and need to be clearly articulated in authentic mathematics educational settings. Secondly, engaging in mathematical problem-solving through CT and programming requires good background in mathematics, algorithmic thinking, and familiarity with the programming language constructs.

The study is still a work in progress. The authors will continue working with undergraduate students and programming courses with applications in mathematics using interventions involving a wide range of participants and whole classes. Future work aims at investigating the following issues.

Firstly, future research will analyze the application of CT in varied, more differentiated, and ill-defined mathematical tasks to deepen the understanding on how CT skills are deployed and applied to solve these kinds of mathematical tasks.

Secondly, future research will focus on more elaborated and in-depth analysis of students' experiences with mixed methods to ensure more reliability and validity of the results. Work will include more participants and a mix of quantitative and qualitative methods to assess the impact of CT on students' learning and uncover what benefits and obstacles different students actually experience in authentic educational settings while thinking computationally to solve mathematical problems, in line with the research literature (Broley et al., 2018).

Thirdly, in terms of the theoretical framework being used in this study, future research work will explore the three-step iterative approach to mathematical problem-solving in more details and depth, in particular the interactions between mathematical thinking, CT, and programming to highlight their communalities and potential differences and suggest changes to the proposed approach. Future research will also look at alternative theoretical frameworks and learning theories used in mathematics education such as Vygotsky's sociocultural theory, Trouche's instrumental approach, Brousseau's theory of didactical situations, or Chevallard's anthropology theory of didactics, but also constructionistic approaches to examine carefully whether these frameworks and theories can be combined and coordinated with the approach presented in this study.

Fourthly, future research should consider usability issues of programming languages since technicalities and familiarity with the language are prerequisites for mathematical problem-solving. Ease of use and understanding of language constructs, interactions, and program feedback of the programming language should be considered in future implementations since these are useful to foster successful implementations of algorithmic solutions to mathematical problem-solving.

Finally, future research will conduct an in-depth examination of how teachers apply CT and programming. This would allow researchers to further probe what teachers view, understand, and interpret as CT and explore how mathematical problem-solving relates to CT and programming from the teacher's point of view.

References

Artigue, M., et al. (2009). Connecting and integrating theoretical frames: The TELMA contribution. *International Journal of Computers for Mathematical Learning, 14*, 217–240.

Bokhove, K., & Drijvers, P. (2010). Digital tools for algebra education: Criteria and evaluation. *International Journal of Mathematics Learning, 15*, 45–62.

Broley, L., Caron, F., & Saint-Aubin, Y. (2018). Levels of programming in mathematical research and university mathematics education. *The International Journal of Research in Undergraduate Mathematics Education, 4*, 38–55.

Buteau, B., Muller, E., Marshall, N., Sacristán, A. I., & Mgombelo, J. (2016). Undergraduate mathematics students appropriating programming as a tool for modelling, simulation, and visualization: A case study. *Digital Experiences in Mathematics Education, 2*, 142–166.

Buteau, C., Muller, E., Mgombelo, J., Sacristán, A. S., & Dreise, K. (2020). Instrumental genesis stages of programming for mathematical work. *Digital Experiences in Mathematics Education, 6*, 367–390.

Buteau, C., et al. (2018). Computational thinking in university mathematics education: A theoretical framework. In A. Weinberg et al. (Eds.), *Proceedings of the 21st annual conference on research in undergraduate mathematics education* (pp. 1171–1179). RUME.

DeJarnette, A. (2019). Students' challenges with symbols and diagrams when using a programming environment in mathematics. *Digital Experiences in Mathematics Education, 5*(1), 36–58.

Filho, P., & Mercat, C. (2018). Teaching computational thinking in classroom environments: A case for unplugged scenario. In V. Gitirana et al. (Eds.), *Proceedings of the re(s)sources 2018 international conference*. ENS de Lyon.

Gueudet, G., et al. (2020). *Programming as an artefact: What do we learn about university students' activity? INDRUM 2020.* Université de Carthage, Université de Montpellier, Cyberspace (virtually from Bizerte), Tunisia. Retrieved from https://hal.archives-ouvertes.fr/hal-03113851/document

Hadjerrouit, S., & Gautestad, H. H. (2019). Evaluating the usefulness of the visualization tool SimReal+ for learning mathematics: A case study at the undergraduate level. In D. Sampson et al. (Eds.), *Learning technologies for transforming large-scale teaching, learning, and assessment* (pp. 71–89). Springer Nature Switzerland AG.

Kotsopoulos, D., Floyd, L., Nelson, V., Makosz, S., & Senger, N. (2019). Mathematical or computational thinking? An early years' perspective. In K. M. Robinson, H. P. Osana, & D. Kotsopoulos (Eds.), *Mathematical learning and cognition in infancy and early childhood: Integrating interdisciplinary research into practice*. Springer.

Kuo, W.-C., & Hsu, T.-C. (2020). Learning computational thinking without a computer: How computational participation happens in a computational thinking board game. *The Asia-Pacific Education Researcher, 29*(1), 67–83.

Lee, I., & Malyn-Smith, J. (2020). Computational thinking integration patterns along the framework defining computational thinking from a disciplinary perspective. *Journal of Science Education and Technology, 29*, 9–18. https://doi.org/10.1007/s10956-019-09802-x

Li, Y., et al. (2020a). On computational thinking and STEM education. *Journal for STEM Education Research, 3*, 147–166.

Li, Y., et al. (2020b). Computational thinking is more about thinking than computing. *Journal for STEM Education Research, 2020*(3), 1–18.

Lie, J., Hauge, I. O., & Meaney, T. J. (2017). Computer programming in the lower secondary classroom: Mathematics learning. *Italian Journal of Educational Technology, 25*(2), 27–35.

Malyn-Smith, J., & Angeli, C. (2020). Computational thinking. In A. Tatnall (Ed.), *Encyclopedia of education and information technologies* (pp. 333–340). Springer Nature Switzerland AG.

Misfeldt, M., & Ejsing-Duun, S. (2015). Learning mathematics through programming: An instrumental approach to potentials and pitfalls. In K. Krainer & N. Vondrová (Eds.), *Proceedings of CERME9* (pp. 2524–2530). Prague, Czech Republic.

Papert, S., & Harel, I. (1991). *Constructionism*. Ablex Publishing.

Patton, M. Q. (2002). *Qualitative research and evaluation methods*. Sage Publications.

Pérez-Marín, D., Hijón-Neira, R., Bacelo, A., & Pizarro, C. (2020). Can computational thinking be improved by using a methodology based on metaphors and scratch to teach computer programming to children? *Computers in Human Behavior, 105*, 105849.

Romero, M., Lepage, A., & Lille, B. (2017). Computational thinking development through creative programming in higher education. *International Journal of Educational Technology in Higher Education, 14*, 1–15.

Santos, S. C., Tedesco, P. A., Borba, M., & Brito, M. (2020). Innovative approaches in teaching programming: A systematic literature review. In *Proceedings of the 12th International Conference On Computer Supported Education* (CSEDU 2020) (Vol. 1, pp. 205–2014). Prague, Czech Republic.

Shodiev, H. (2015). Computational thinking and simulation in teaching science and mathematics. In M. G. Cojocaru et al. (Eds.), *Interdisciplinary topics in applied mathematics* (Springer proceedings in mathematics and statistics) (Vol. 117, pp. 405–410).

Shute, V. J., Sun, C., & Asbell-Clarke, J. (2017). Demystifying computational thinking. *Educational Research Review, 22*, 142–158.

Topalli, D., & Cagiltay, N. E. (2018). Improving programming skills in engineering education through problem-based game projects with scratch. *Computers and Education, 120*, 64–74.

Weintrop, D., et al. (2016). Defining computational thinking for mathematics and science classrooms. *Journal of Science Education and Technology, 25*, 127–147.

Wing, J. M. (2008). Computational thinking and thinking about computing. *Philosophical Transactions of the Royal Society A, 366*(1881), 3717–3725.

Wing, J. M. (2014). Computational thinking benefits society. *Social issues in computing,* 40th Anniversary Blog, University of Toronto. Retrieved from http://socialissues.cs.toronto.edu/index.html%3Fp=279.html

Chapter 12
Summary Writing with Graphic Organizers in Web-Based Investigative Learning

Oriko Harada and Akihiro Kashihara

12.1 Introduction

Building knowledge online has become an important skill as mobile devices and personal computers provide people with easier access to information on the Web. The Web is one of the main resources for people to collect information for various purposes (Kammerer, Brand-Gruwel, & Jarodzka, 2018). In particular, it has been reported that students often use online information for completing most of their assignments (Julien & Barker, 2009). Gross and Latham (2013) noted that students found various difficulties in assessing whether the information was valid and reliable while gathering resources and also in organizing and synthesizing the information for building their knowledge.

Searching as learning (SAL) is one recent approach to the way learners engage with search systems (Hansen & Rieh, 2016). This is characterized by viewing searching as not an isolated activity where learners enter a single query to conduct a simple information retrieval task, but as a discovery process for building a body of knowledge (Rieh, Collins-Thompson, Hansen, & Lee, 2016). Rieh, Gwizdka, Freund, and Collins-Thompson (2014) differentiate these two modes of searching as "look-up tasks" and "exploratory search." They argue that current systems are optimized for the former, which involve minimizing interaction times and lowering user effort. They also argue that more research and support is required for the latter involving a richer engagement with the information retrieved.

Information seeking on the Web consists of multiple processes, such as selecting suitable Web resources and navigating between Web pages, and various works have

O. Harada · A. Kashihara (✉)
Graduate School of Informatics and Engineering, The University of Electro-Communications, Chōfu, Tokyo, Japan
e-mail: akihiro.kashihara@inf.uec.ac.jp

D. Ifenthaler et al. (eds.), *Orchestration of Learning Environments in the Digital World*, Cognition and Exploratory Learning in the Digital Age,
https://doi.org/10.1007/978-3-030-90944-4_12

been conducted to provide support. For example, Nishihara, Sunayama, and Yachida (2006) aimed to support learners in selecting resources according to their knowledge about a certain topic. To this end, they developed an application, which extracts keywords from Web resources and provides learners with a difficulty in learning each resource and the order of reading the resources to understand the topic. Aramaki, Tsutsui, and Takada (2016) aimed at supporting elementary school students to internalize what they read from resources by providing a tablet application where students can save images from resources and a note attached to each image. Our aim, on the other hand, is to deepen and widen the learners' investigation on the Web. To do so, we have defined a learning model where learners seek information with not only a single query but multiple search queries (Kashihara & Akiyama, 2017).

We have defined *Web-based investigative learning (WBIL for short)* as a form of learning where learners use a search engine to investigate any question with Web resources to construct knowledge. In WBIL, learners are provided with a learning model and a cognitive tool where they can create a learning scenario visualized as a tree with each search query as its node and the progress of one search query to another as links between the nodes. The users of WBIL are beginners in searching information online, but basic knowledge of using computers and reading skills to comprehend online articles are presupposed. In practice, university students would be the main users, and our past experiments have been conducted on undergraduate and graduate students (Kashihara & Akiyama, 2017; Morishita, Kashihara, Ota, & Hasegawa, 2020). The learning model and iLSB have been then used not only for investigation online but also for reflecting on, organizing, and presenting knowledge constructed from Web resources. The current work focuses on the knowledge presentation process, where learners are expected to write a paper reporting findings from the investigative learning on the Web.

Although information gathered/constructed with iLSB are useful for learners themselves in reflecting upon their own investigative learning process, it is difficult to understand other learners' investigation obtained with iLSB. In order for learners to share their learning outcomes with peers or instructors, they are required to convert the information stored in iLSB into a paper. On the other hand, the paper should include the contents learners investigated and contain a substantial amount of information for readers of the paper. Therefore, a summary for the paper is necessary as an advance organizer for learners to effectively share what they investigated. As there has been no established method for learners of WBIL to write a summary, which is not a simple task, we propose a method to support summary writing in WBIL (Harada & Kashihara, 2020). This method uses graphic organizers, which represent the text structure of a summary. We have also reported a case study of the method that we conducted with 16 participants (Harada & Kashihara, 2021). The results of the study suggest that it allows learners to write an exhaustive summary with lower cognitive load.

12.2 Web-Based Investigative Learning

12.2.1 Learning Model

In this paper, we introduce a learning model for Web-based investigative learning (WBIL), which consists of the following two processes: (1) knowledge construction process and (2) knowledge linearization process. Each process further has several learning phases. The learning model designed in our previous work (Hagiwara, Kashihara, Hasegawa, Ota, & Takaoka, 2019; Kashihara & Akiyama, 2017) represents the process (1). In the subsequent work (Morishita et al., 2020) and this paper, we have been attempting to extend this model to include the process (2).

12.2.1.1 Knowledge Construction Process

In the knowledge construction process, learners are expected to investigate a question called initial question, which is often represented such as "What is X?" They are then expected to use a search engine with a keyword (called q-keyword) representing the question to explore Web resources and to build comprehensive knowledge. Such knowledge construction process is composed of the following three phases (Kashihara & Akiyama, 2017): (1-a) search for Web resources, (1-b) navigational learning, and (1-c) question expansion.

In (1-a), learners are expected to search Web resources with the initial q-keyword. Using the searched results, they are next expected to compare multiple resources to find out Web pages suitable for investigating the initial question. In (1-b), learners are expected to navigate and learn the Web pages and to combine the contents learned at each page and construct their knowledge. This phase involves extracting keywords from the Web pages to make connections among these keywords and constructing their knowledge for the corresponding q-keyword. In (1-c), learners are expected to find out sub-questions to be further investigated from knowledge constructed. In other words, learners are allowed to expand the initial question into sub-questions. They are expected to cyclically repeat the three phases for each sub-question until no further question expansion occurs. The question expansion results in a tree structure called question tree as shown in Fig. 12.1, which represents a learning scenario. The root node corresponds to the initial q-keyword, and the other nodes correspond to q-keywords representing the corresponding sub-questions.

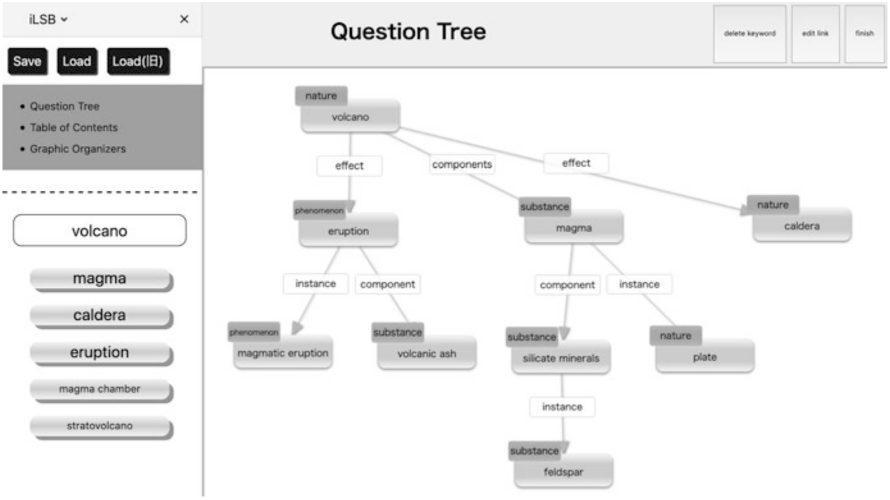


Fig. 12.1 An example of question tree

12.2.1.2 Knowledge Linearization Process

In the knowledge linearization process, learners are expected to write a paper from their question tree created, which includes findings in their investigation, and its summary. In the paper writing, first, they are allowed to use their question tree to linearize their knowledge constructed. We have provided a scaffold as iLSB function for learners to generate a table of contents (TOC) from a question tree (Morishita et al., 2020). Figure 12.2 shows an example of TOC constructed from the question tree shown in Fig. 12.1. TOC has a hierarchical structure, in which each node corresponds to q-keyword in the question tree. The nodes represent chapters, sections, subsections, etc. of the paper to be written. The relationships between a chapter and its sections and the ones between a section and its subsections are represented as links attached in the knowledge construction process with iLSB. Each link has one of the 14 attributes such as "class," "instance," and "effect." How to create TOC is beyond the scope of this work (Morishita et al., 2020).

Second, the summary writing has the following two phases: (2-a) text structuring with graphic organizers and (2-b) text writing. In (2-a), learners are expected to use graphic organizers to structure a text to be included in the summary. They are allowed to reorganize the nodes in TOC with graphic organizers to visually represent the text structure to be embedded in the summary of their paper. They are then expected to compose sentences from the text structure with graphic organizers.

In the text structuring phase, learners are induced to make connections between q-keywords, which allows them to overview and rethink their knowledge not with the question tree but with the text structure. In addition, learners could share the summary written as for what they learned with their peers and instructors. The instructors could use it to evaluate the learners' investigative learning for the initial question.

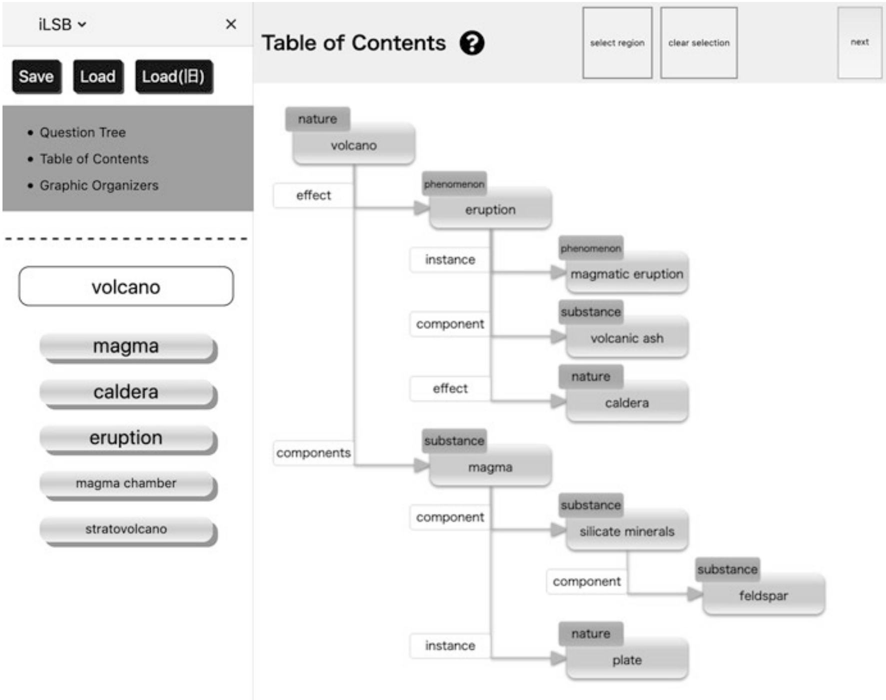


Fig. 12.2 An example of TOC

Through the paper writing process, learners could reflect on their knowledge constructed and write a summary of the paper to report on their investigation outcome. In this work, we address the issue how to help learners build a summary for the paper with graphical organizers.

12.2.2 iLSB

iLSB aims to scaffold WBIL as modeled, which functions as a Firefox add-on. In order to scaffold the knowledge construction process, iLSB provides learners with a search engine for Web resources, a keyword repository for collecting keywords to be extracted from the resources during navigational learning to represent knowledge constructed, and a question tree for question expansion.

iLSB also provides a scaffold for the knowledge linearization process, which allows learners to create and edit a TOC from the question tree constructed in the knowledge construction process. iLSB first lists nodes (represented as q-keywords) in the question tree as candidates for chapters and sections. Learners are expected to select some nodes as TOC chapters from the candidate list and then to select some nodes as sections for each chapter. They are next expected to make part-of relationships between the TOC chapters and their section nodes and also define subsections in each section. They are finally expected to sort the sections/subsections/chapters to define the order of nodes that is best suited for explaining the initial q-keyword.

In addition to the scaffold for creating and editing TOC, we have developed a scaffold for allowing learners to write a summary with graphical organizers.

12.3 Summary Writing for WBIL

Let us here explain how to scaffold the summary writing for WBIL with iLSB.

12.3.1 Framework

In this work, we assume that learners use iLSB for conducting WBIL as modeled and create their own question tree and TOC. They are then expected to write a summary involving their investigative learning outcome. In order to scaffold the summary writing, iLSB first generates a text structure involving q-keywords from TOC, which are represented with graphic organizers. Figure 12.3 shows two examples of text structure, each of which includes some nodes representing q-keywords. Graphic organizers have five basic structure types: description, sequence and order, cause and effect, problem and solution, and compare and contrast (Roehling, Hebert, Nelson, & Bohaty, 2017). iLSB then allows learners to add two types of annotations to the generated text structure. These annotations enable them to supplement the contents of nodes and the relationships between nodes. They are finally expected to use the text structures with annotation to write the text of the summary.

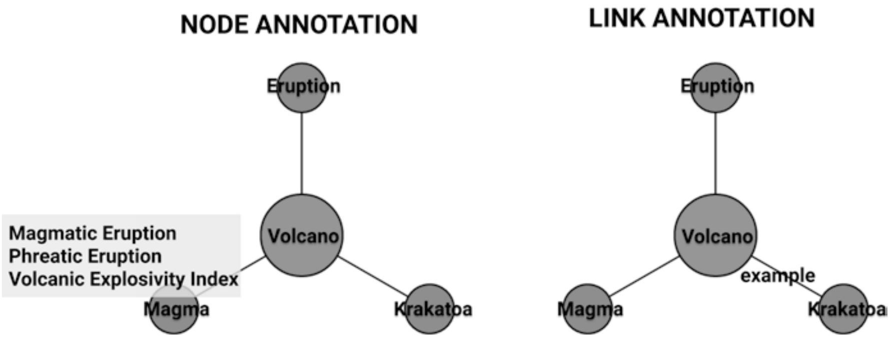


Fig. 12.3 Text structures with node and link annotations

12.3.2 Representing Text Structure with Graphic Organizers

It has been widely used in conventional education to use graphic organizers to help learners in reading or writing a text (Roehling et al., 2017). In this work, we apply this method to the summary writing for WBIL. Graphic organizers visually represent basic structures of expository text. The left structure in Fig. 12.3 shows an example of graphic organizer for description about the subject of "Volcano," in which keywords "Krakatoa," "Eruption," and "Magma" are used for describing the subject. It allows learners to compose a text such as "Volcano has eruption and contains magma. An example of volcano is Krakatoa."

Learners are also allowed to annotate nodes and links in text structure represented with graphic organizers. Node annotations allow learners to add keywords to any node, which supplement it. The left structure in Fig. 12.3 shows the node "Eruption" has annotations including "Magmatic Eruption," "Phreatic Eruption," and "Volcanic Explosivity Index." Link annotations also allow learners to add keywords to any link. In Fig. 12.3, the link between "Volcano" and "Krakatoa" is annotated with "example." These annotations allow learners to refer to the annotated keywords to write sentences for summary writing.

12.3.3 Text Structuring of Summary in iLSB

iLSB helps learners build a text structure with graphic organizers in the following four steps. First, iLSB shows TOC and allows learners to review the nodes, links, and hierarchical order. Second, iLSB automatically generates a text structure for summary from TOC to present.

Let us explain the procedure for generating a text structure with graphic organizers with an example of TOC shown in Fig. 12.2. First of all, iLSB groups q-keywords included in TOC into the ones for chapters/sections/subsections. In this example, the groups are {volcano, eruption, magma} as chapters, {eruption, magmatic eruption, volcanic ash, caldera} and {magma, silicate minerals, plate} as sections, and {silicate minerals, feldspar} as subsections. Following the relationships (attributes attached) between the q-keywords in each group, iLSB second selects a basic structure of text.

Table 12.1 shows the correspondence of the attributes attached between the q-keywords in TOC to basic structure of text. If there are multiple attributes in the group, iLSB splits the group as to each basic structure. In the group {eruption, magmatic eruption, volcanic ash, caldera}, for example, the attribute attached between "eruption" and "caldera" is "effect," the one between "eruption" and "magmatic eruption" is "instance," and the one between "eruption" and "volcanic ash" is "component." The attribute "effect" corresponds to the basic structure "cause and effect," while the attributes "instance" and "component" correspond to "description." The group will be accordingly split into the two subgroups {eruption, caldera} and

Table 12.1 Correspondence of attributes to basic structure of text

Attributes attached between q-keywords in TOC	Basic structure of text
Class, instance, component, structure, characteristic	Description
Origin, background	Sequence and order
Cause, effect, principle, action	Cause and effect
Measure	Problem and solution
Similar concept, contrasting concept	Compare and contrast

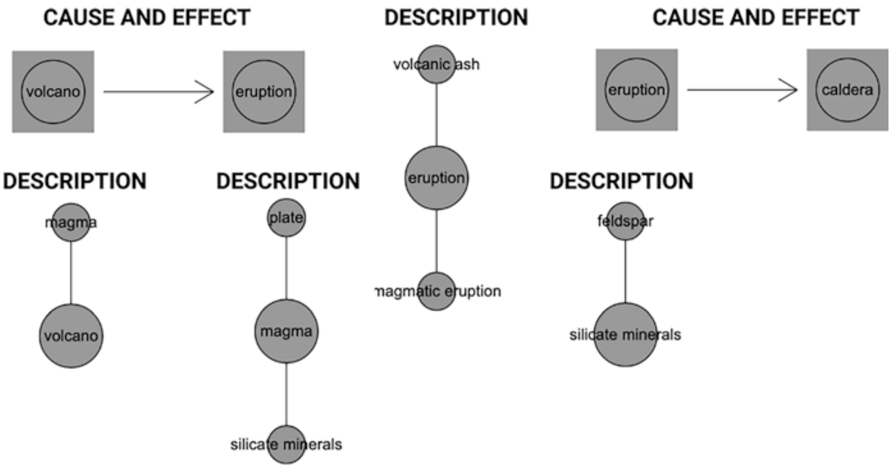


Fig. 12.4 An example of text structure with graphic organizers

{eruption, magmatic eruption, volcanic ash}. As a result, iLSB generates the text structure embedding the six basic structures as shown in Fig. 12.4 from TOC in Fig. 12.2.

When a text structure generated is presented to learners, they are third expected to understand the basic structures to be used in their summary and the q-keywords included. Fourth, learners are expected to add node and link annotations to the basic structures. The keywords used for node annotations are selected from the keyword repository in iLSB, which stores keywords representing the contents learned about the node. As for link annotations, learners could use their own keywords.

iLSB finally provides learners with the generated text structure. By clicking on any node, learners could see the node annotations. Link annotations are visible without any operation. After building a text structure, learners are allowed to start writing a summary.

12.4 Case Study

Let us next describe the case study we have conducted whose purpose was to assess the effectiveness of our method in helping learners write a summary for WBIL. In particular, we set the following perspectives for evaluation: (1) exhaustiveness of summary with respect to the corresponding TOC and (2) cognitive load of summary writing.

12.4.1 Methods

Participants were 16 undergraduate and graduate students at the University of Electro-Communications, Japan. As their major was informatics and engineering, they could be assumed to have sufficient computer literacy and would be suited as WBIL users.

We also planned the study within 180 min at the maximum. We considered a within-subjects design too exhausting for the participants. Moreover, potential learning effects would confound any differences found in the study. As such, participants were equally divided into either the experimental group (eight participants) or the control group (eight participants) under a between-subjects design.

The procedure of the study was as follows. The participants were first required to investigate the initial q-keyword “Tax” (30–60 min) with iLSB. They were second required to create TOC for the paper with iLSB (20–30 min) and then to finally write a summary (30–60 min). In the summary writing, the experimental group were allowed to use graphic organizers to build a text structure from their TOC (15–30 min) and to view it while writing the summary. The control group was allowed to write a summary only with the TOC (without graphic organizers). In addition, we also conducted a reviewer assessment where two reviewers who were familiar with WBIL measured how well a summary written by each participant reflected the structure and contents of TOC.

After summary writing, we collected and analyzed data related to the two evaluation perspectives mentioned above: exhaustiveness of summary and cognitive load of summary writing. For the former, we analyzed the number of q-keywords and letters in summaries written. We also conducted a questionnaire (questionnaire #1) to ask all participants how exhaustive they were able to write their summaries. For the latter, we analyzed time required to write one letter in summary. We also asked all participants how difficult is writing a summary in the same questionnaire. The experimental group, in addition, were asked about iLSB use in creating text structures and general impression on the method proposed by means of another questionnaire (questionnaire #2).

12.4.2 Results

12.4.2.1 Analysis of Summary

Table 12.2 shows the results of analyzing the summary text with the collected data as for (1) exhaustiveness of summary, (2) length of summary, and (3) seconds spent for summary writing.

As for the exhaustiveness of summary, which was calculated as the ratio of q-keywords in summary to q-keywords in TOC, first, the control group covered 91% (SD, 0.12) of the TOC q-keywords in the summary on average. The experimental group also covered 96% (SD, 0.06). As for the length of summary, second, we collected the total number of letters in the summary as well as the number of letters per q-keyword. The control group had 332 letters (SD, 99.71) and 29.43 letters per q-keyword (SD, 13.96) on average. The experimental group also had 558 letters (SD, 343.74) and 42.97 letters per keyword (SD, 12.08). As for seconds spent for writing the summary, the control group had 4.05 s per letter (SD, 1.77) on average, and the experimental group had 2.97 s per letter (SD, 0.99).

From these results, we could see some trends where the experimental group tended to cover more of the q-keywords in TOC, take less time to write their summaries, and write more as for each q-keyword, although there were no statistically significant differences between the two groups.

12.4.2.2 Reviewer Assessment

In the reviewer assessment, two reviewers read each summary generated by each participant and then drew TOC which they could assume that the summary was created from. TOC drawn was then compared to the original TOC created by the participant, and the tree edit distance (TED) (Zhang & Shasha, 1989) between these TOCs was calculated. This assessment aimed to measure how well the summary reflected the structure and contents of TOC.

The correlation between the two reviewers' TEDs was calculated as 0.71 that seemed high. We have accordingly viewed the average of their TEDs as the distance between the reviewers' TOC and participant's TOC. The average TED was 8.88

Table 12.2 Results of summary analysis

Data	Control group		Experimental group	
	Average	SD	Average	SD
Q-keywords in summary/q-keywords in TOC	91%	0.12	96%	0.06
Number of letters	332 letters	99.71	558 letters	343.74
Number of letters/q-keyword	29.43 letters/keyword	13.96	42.97 letters/keyword	12.08
Seconds spent/letter	4.05 s/letter	1.77	2.97 s/letter	0.99

(SD, 2.98) for the control group and 6.38 (SD, 3.31) for the experimental group. The experimental group has a shorter TED than the control group, although there was no statistical significance. This suggests the experimental group tends to write more suitable summary.

In addition, some participants in the control group seemed to possess a high writing skill since they had shorter TEDs than the average TED of the experimental group. We accordingly divided the control group into two subgroups with lower skill and higher skill and compared the experiment group and the control group with lower skill involving three participants who had longer TEDs than the TED average of the control group. The results showed the subgroup's TED average was 11.67 (SD, 2.83), which made a larger difference between the groups.

12.4.2.3 Questionnaire #1

Table 12.3 shows the questionnaire #1, which was used for both groups. It included three questions on a five-point Likert scale. The average scores for each group are shown in Table 12.3.

The first question (Q1-1) was related to the subjective difficulty of summary writing, where 1 on the Likert scale indicated "very easy" and 5 indicated "very difficult." The average score of the control group was 2.50 (SD, 0.93), and the one of the experimental group was 2.38 (SD, 0.74). The second question (Q1-2) was also related to the exhaustiveness of the written summary from the participant's point of view, where 1 indicated "not exhaustive at all" and 5 indicated "very exhaustive." The average score of the control group was 4.13 (SD, 0.83), and the one of the experimental group was 4.25 (SD, 0.46). The third question (Q1-3) was related to the subjective effect of summary writing in gaining an overview of the contents learned in their investigation, where 1 indicated "not effective at all" and 5 indicated "very effective." The average score of the control group was 4.38 (SD, 0.52), and one of the experimental group was 4.50 (SD, 0.53).

Table 12.3 Results of questionnaire #1

Question number	Question	Control group average	Experimental group average
Q1-1	How difficult was the summary writing?	2.50 (SD, 0.93)	2.38 (SD, 0.74)
Q1-2	How much were you able to cover what you investigated in the summary?	4.13 (SD, 0.83)	4.25 (SD, 0.46)
Q1-3	Does the summary writing allow you to gain an overview of what you investigated?	4.38 (SD, 0.52)	4.50 (SD, 0.53)

12.4.2.4 Questionnaire #2

Tables 12.4 and 12.5 show the results of the questionnaire #2, which was used only for the experimental group. This questionnaire consisted of the following three sections: (A) ease of using graphic organizers, (B) effect of creating text structures with graphic organizers, and (C) process of summary writing with graphic organizers. The questions given in this questionnaire were either answered with a five-point Likert scale or open-ended way. Tables 12.4 and 12.5 show the average scores for the questions. The questions answered with open-ended way are marked with "-" in the tables.

Table 12.4 shows the results concerning the use of graphic organizers. In this table, 1 on the Likert scale indicated "very difficult," and 5 indicated "very easy." The ease of using graphic organizers scored an average of 4.1 (SD, 0.83). The open-ended answers to Q2A-2 suggested that some participants had a difficulty in entering link annotations, viewing node annotations or basic structures represented with graphic organizers, and moving between basic structures generated to enter annotations. One participant commented that the basic structures generated by iLSB were different from what the participant predicted and that this led to some difficulty particularly in adding annotations. The ease of entering node and link annotations averaged 4.1 (SD, 0.99) and 3.8 (SD, 1.58). As for the ease of grasping basic structures generated, the average was 4.1 (SD, 0.99).

Table 12.5 shows the results as for learning effect of graphic organizers. As for Q2B-1 and Q2B-2, 1 on the Likert scale indicated "not helpful at all," and 5 indicated "very helpful." The average score of Q2B-1 was first 3.8 (SD, 1.04)

Table 12.4 Results on the use of graphic organizers (questionnaire #2)

Question number	Question	Average	SD
Q2A-1	Was it easy to use iLSB to create text structure with graphic organizers?	4.1	0.83
Q2A-2	(continued from Q2A-1) what were the difficulties?	–	–
Q2A-3	Was it easy to enter node annotations?	4.1	0.99
Q2A-4	Was it easy to enter link annotations?	3.8	1.58
Q2A-5	Was it easy to visually grasp basic structures represented with graphic organizers when writing the summary?	4.1	0.99

Table 12.5 Results on learning effect of graphic organizers (questionnaire #2)

Question number	Question	Average	SD
Q2B-1	Did graphic organizers help you organize information in TOC?	3.8	1.04
Q2B-2	Did graphic organizers help you organize information from your investigation?	4.1	1.04
Q2B-3	How well did you understand text structures and other information represented with graphic organizers?	4.0	0.93
Q2B-4	What were the factors of graphic organizers that made it difficult to use?	–	–

Table 12.6 Results on the process of summary writing with GOs (questionnaire #2)

Question number	Question	Average	SD
Q2C-1	How useful were the node annotations for summary writing?	4.6	0.52
Q2C-2	How useful were the link annotations for summary writing?	4.1	0.99
Q2C-3	How effective were graphical organizers in writing an exhaustive summary?	3.9	0.83
Q2C-4	What were the factors of graphical organizers that made it easy to write a text?	–	–
Q2C-5	Was it difficult to write a text from graphical organizers?	1.9	0.64
Q2C-6	What were the factors of graphical organizers that made it difficult to write a text?	–	–
Q2C-7	Please write any other comments on graphical organizers and summary writing	–	–

concerning the effectiveness of graphic organizers in organizing information in TOC. The result from Q2B-2 showed the effectiveness in organizing the information such as question tree and TOC obtained from iLSB, and the average score was 4.1 (SD, 1.04). Q2B-3 aimed to measure how well the participants understood the text structure represented with graphical organizers on a Likert scale with 1 indicating "could not understand at all" and 5 indicating "understood very well." The average score for this question was 4.0 (SD, 0.93). Finally, Q2B-4 aimed to detect factors of graphical organizers to be improved as future work. The participants raised the following points as difficulties in using graphical organizers:

– Inability to change basic structures generated with graphical organizers.
– Necessity of graphical organizers to be one of the five types.
– Necessity of checking basic structure represented with each graphical organizer.

Table 12.6 shows the results concerning the process of summary writing with graphic organizers. In Q2C-1 and Q2C-2, the participants were asked how much they used the information in node and link annotations for summary writing. The average scores on a Likert scale with 1 indicating "not useful at all" and indicating 5 "very useful" were 4.6 (SD, 0.52) for the node annotations and 4.1 (SD, 0.99) for link annotations.

In Q2C-3, they were also asked the perceived effectiveness of graphic organizers in writing an exhaustive summary. They answered on a Likert scale with 1 indicating "not effective at all" and 5 indicating "very effective." The average score was 3.9 (SD, 0.83). In Q2C-4, they were also asked any other factors of graphical organizers that made it easy to write a text, and they listed the following:

- No need to take notes for writing.
- Visualization of the thought process.
- Visualization of the relationships and relevance between questions investigated.
- Ease of ordering the questions in the summary.
- Ability to gain an overview of the questions investigated.

In Q2C-5, the participants were asked the difficulty of writing a text from graphical organizers where they answered on a Likert scale with 1 indicating “very easy” and 5 indicating “very difficult.” The average score was 1.9 (SD, 0.64). In Q2C-6, they were also asked any factors, which made the above process difficult. Some participants commented that there were some difficulties in writing sentences from keywords as nodes in the text structure and in recalling what they learned from the investigation. In Q2C-7, they were asked some comments regarding summary writing with graphical organizers. We could observe some valuable comments such as:

– The ability to visualize, record, and organize information obtained from investigative learning with iLSB in addition to own ideas was perceived to be an advantage of graphical organizers.
– Graphical organizers made summary writing easier.

12.4.3 Discussion

Let us finally discuss the results of the case study from the two viewpoints, which are (1) exhaustive summary writing and (2) cognitive load of summary writing.

First, the experimental group was expected to write a longer and more exhaustive summary in less time (easier) compared to the control group. As expected, the experimental group performed well on the exhaustiveness of summary, which covered 96% of the q-keywords in TOC on average. On the other hand, the control group also covered 91%, which seems high. In this study, we could not have any significant differences between the experimental and control groups. By observing individual data, this can be partly attributed to the participants in the control group who have had a high skill in organizing and summarizing information. The reviewer assessments also revealed that compared to the subgroup of the control group with a lower skill, the experimental group’s summary reflected TOC better.

The results from questionnaire #1 also suggest that the control group already had a high writing and summarizing skill. From Q1–1, we can observe that the perceived difficulty in summary writing was the same for both groups even though the control group had no support for it. This suggests that there are some learners who do not need any support for summary writing in WBIL. We think such learners are not target users of the method proposed in this paper. As such, we can conclude that the proposed method allows learners who have a lower skill in summary writing to reorganize an exhaustive summary from the information investigated in WBIL.

From the results of summary analysis shown in Table 12.2, second, graphic organizers allow learners to write a summary more efficiently, which suggest it contributes to reducing cognitive load of summary writing. Questionnaire #2 shown in Table 12.4 also suggests that the participants can use graphic organizers to create a text structure in an easy way.

From these results and suggestions, we can conclude that using graphic organizers for summary writing is partially successful from the two viewpoints. However, we need a more detailed study with a larger number of participants as future work.

12.5 Conclusion

Following the Web-based investigative learning (WBIL) model we designed, iLSB allows learners to effectively investigate information on the Web and construct a question tree. In this paper, we focus on summary writing for WBIL with iLSB, which is an essential process for learners to share their learning outcomes/findings to their peers or instructors.

We have accordingly designed a method for helping learners build text structure with graphic organizers from TOC to be created from the question tree in WBIL. This method allows them to write a summary for their investigation in WBIL with five types of graphic organizers, which represent the five basic structures (description, sequence and order, cause and effect, compare and contrast, and problem and solution). Learners are allowed to reorganize nodes included in TOC by means of these graphic organizers to build a text structure of the summary.

We have also conducted a case study to evaluate the method. The results suggest that text structure generated with graphic organizers functions effectively as a scaffold for writing an exhaustive summary with less effort. Although we could not obtain any significant differences between the control and experimental groups, the questionnaire results from the experimental group showed an overall positive attitude toward the method. The graphic organizers visualize the order and relationships among information to be written in the summary to facilitate learners' summary writing. In addition, the summary writing allows learners to gain an overview perspective of what they learned in their investigation. Along with the data obtained from the summary analysis, we can conclude that graphic organizers are able to decrease learners' cognitive load in writing a summary and making its contents exhaustive.

In the future, we plan to conduct a detailed case study with more control over participants' skills.

Acknowledgments The work presented here is partially supported by JSPS KAKENHI Grant Number 17H01992 and 18H01053.

References

Aramaki, M., Tsutsui, K. & Takada, H. (2016). An investigative learning support system on the web (in Japanese). In *Proceedings of the 78th National Convention of IPSJ* (Vol. 1, pp. 673–674).

Gross, M., & Latham, D. (2013). Addressing below proficient information literacy skills: Evaluating the efficacy of an evidence-based educational intervention. *Library and Information Science Research, 35*(3), 181–190.

Hagiwara, H., Kashihara, A., Hasegawa, H., Ota, K., & Takaoka, R (2019). Adaptive recommendation for question decomposition in web-based investigative learning. In *The 2019 IEEE International Conference on Teaching, Assessment, and Learning for Engineering (*TALE2019*)* (pp. 750–757).

Hansen, P., & Rieh, S. Y. (2016). Recent advances on searching as learning: An introduction to the special issue. *Journal of Information Science, 42*(1), 3–6.

Harada O. & Kashihara, A. (2020). A graphic organizer for supporting summary writing in web-based investigative learning. In *The 17th International Conference on Cognition and Exploratory Learning in Digital Age (CELDA 2020)* (pp. 257–264).

Harada O. & Kashihara, A. (2021). A case study for summary writing with a graphic organizer in web-based investigative learning. In *14th IADIS International Conference on Information Systems (IS2021)* (pp. 113–120).

Julien, H., & Barker, S. (2009). How high-school students find and evaluate scientific information: A basis for information literacy skills development. *Library and Information Science Research, 31*(1), 12–17.

Kammerer, Y., Brand-Gruwel, S., & Jarodzka, H. (2018). The future of learning by searching the web: Mobile, social, and multimodal.

Kashihara, A., & Akiyama, N. (2017). Learning scenario creation for promoting investigative learning on the Web. *The Journal of Information and Systems in Education, 15*(1), 62–72.

Morishita, N., Kashihara, A., Ota, K., & Hasegawa, S. (2020). Reconstructing question tree in web-based investigative learning (in Japanese). In *44th Annual Conference of Japanese* Society for Information and Systems in Education, Hamamatsu, Japan.

Nishihara, Y., Sunayama, W., & Yachida, M. (2006). Information understanding support by using informational complexity and continuity (in Japanese). *IEICE Transactions on Information and Systems, 89*(9), 1963–1975.

Rieh, S. Y., Collins-Thompson, K., Hansen, P., & Lee, H. J. (2016). Towards searching as a learning process: A review of current perspectives and future directions. *Journal of Information Science, 42*(1), 19–34.

Rieh, S. Y., Gwizdka, J., Freund, L., & Collins-Thompson, K. (2014). Searching as learning: Novel measures for information interaction research. *Proceedings of the American Society for Information Science and Technology, 51*(1), 1–4.

Roehling, J., Hebert, M., Nelson, J., & Bohaty, J. (2017). Text structure strategies for improving expository reading comprehension. *The Reading Teacher, 71*(1), 71–82.

Zhang, K., & Shasha, D. (1989). Simple fast algorithms for the editing distance between trees and related problems. *SIAM Journal on Computing, 18*(6), 1245–1262.

Chapter 13
Interaction Preferences in Digital Learning Environments: Does Gender and Achievement Matter?

Muhittin Sahin and Dirk Ifenthaler

13.1 Introduction

Recent advances in information and communication technologies (ICT) provide multiple opportunities to support learners, educators, and digital learning environments. One goal of using ICT in the field of education is to develop and improve learning processes without time and space constraints (Ji, Park, Jo, & Lim, 2016). In this context, many digital learning environments have been developed and implemented in higher education institutions around the world. A major criticism brought to these digital learning environments was that the individual learning activities cannot be monitored consistently (Graf, Kinshuk, & Liu, 2009). Since then, many approaches have been used to understand learners' learning profiles and to predict their future development based on such digital educational data and assessments (Ifenthaler, Pirnay-Dummer, & Seel, 2010). Recent advancements of educational data mining and learning analytics allow a precise tracking of learner activities, i.e., trace data in digital learning environments (Schumacher, Klasen, & Ifenthaler, 2019; Şahin, Keskin, Özgür, & Yurdugül, 2017). Still, recording only trace data does not support a holistic perspective toward learning with regard to learners' context and background (Ifenthaler & Widanapathirana, 2014).

M. Sahin (✉)
Ege University, Izmir, Turkey

Learning, Design and Technology, University of Mannheim, Mannheim, Germany
e-mail: musahin@mail.uni-mannheim.de; muhittin.sahin@ege.edu.tr

D. Ifenthaler
Learning, Design and Technology, University of Mannheim, Mannheim, Germany

UNESCO Deputy Chair on Data Analytics in Higher Education Learning and Teaching, Curtin University, Perth, WA, Australia
e-mail: dirk@ifenthaler.info

D. Ifenthaler et al. (eds.), *Orchestration of Learning Environments in the Digital World*, Cognition and Exploratory Learning in the Digital Age,
https://doi.org/10.1007/978-3-030-90944-4_13

One strategy of educational data mining is the identification and classification of interaction profiles and patterns of learners in order to optimize digital learning environments within the context of learner needs and preferences. As such, it is possible to offer more appropriate and customizable learning environments to the learners. Successful applications have been studied with a focus on learners' navigation profiles (Blagojević & Milosević, 2013; Graf, Kinshuk, & Liu, 2009), academic achievements (Jo, Yu, Lee, & Kim, 2015; Macfadyen & Dawson, 2010), study success (Ifenthaler & Yau 2020), or motivation sources and cognitive strategies (Keskin & Yurdugül, 2019; Schumacher & Ifenthaler, 2018; Şahin et al., 2017).

Within the scope of this research, the interaction preferences of students in a digital learning environment were examined based on their gender and achievement situation. In addition, the interaction preferences of the students were examined according to different phases of learning during a 12-week semester at a higher education institution.

13.2 Background

13.2.1 Learning Management Systems and Learners' Preferences

In recent years, the education paradigm has shifted from one size fits all to an adaptive and customizable approach (Brusilovsky & Henze, 2007). In this context, it is also seen that digital learning environments are used extensively by higher education institutions. Especially, learning management systems (LMSs) are used remarkably (Brown, Dehoney, & Millichap, 2015). LMS is a software designed to help instructors in order to achieve their pedagogical goals by presenting learning content to the students and managing users and rights on the platform (Ifenthaler, 2012; Machado & Tao, 2007). LMSs are used to manage the learning process and learning content (Ji et al., 2016; Ifenthaler & Pirnay-Dummer, 2011) and include several opportunities for stakeholders such as presenting the learning content, conducting exams, collecting assignments, and documenting students' learning process. One of the key features of such digital learning environments is the collection of various interaction types. Interaction has become even more valuable, especially with the emergence of new distance education programs that promote interaction at various levels and types (Brown & Long, 2006). It is possible to consider these types of interactions mainly as learner-content, learner-instructor, and learner-learner interactions, which are put forth by Moore (1989). In addition, recent developments focus on the learner-assessment interaction type (Şahin et al., 2017). Within the scope of this research, the LMS interactions of the learners are defined as (a) learner-content, (b) learner-discussion (learner-learner), (c) learner-tutorial, (d) learner-video, and (e) learner-assessment.

Data of learners' interactions are stored unstructured in LMS environments (Šimić, Gašević, & Devedžić, 2004). This log data may contain multiple ways for determining learning profiles and patterns of learners. The learners' learning profile must be determined in order to support the learning process and present appropriate artifacts and actions in a digital learning environment (Ifenthaler, 2015). Current studies focusing on learning profiles of learners in digital learning environments focus on log data (Keskin, Şahin, & Yurdugül, 2019; Blagojević & Milosević, 2013; Macfadyen & Dawson, 2010) or self-report data (Northrup, 2009; Rhode, 2009). In this research, interaction preferences of the students were examined based on log data and evaluation data collected through digital instruments administered through a learning analytics system. The interactions of students with the learning environment were considered as the number of clicks on the themes and the total frequencies were used in the standardized residuals analysis. Detailed information about the analysis is presented in the method section.

13.2.2 Research Questions

This chapter focusses on interaction preferences of students in a digital learning environment with a specific focus on the influence of their gender and achievement. In addition, the learning experiences of the students were divided into three periods of study during the course of the semester. Accordingly, six research questions have been investigated:

1. Do the interaction preferences differ between females and males in the first study period?
2. Do the interaction preferences differ between females and males in the second study period?
3. Do the interaction preferences differ between females and males in the third study period?
4. Do the interaction preferences differ between successful and unsuccessful students in the first study period?
5. Do the interaction preferences differ between successful and unsuccessful students in the second study period?
6. Do the interaction preferences differ between successful and unsuccessful students in the third study period?

13.3 Method

13.3.1 Context and Participants

The study followed an evidence-centered design approach (Behrens, Mislevy, Dicerbo, & Levy, 2012). Student and evaluation data were collected while implementing the learning analytics system LeAP (Learning Analytics Profile; Klasen & Ifenthaler, 2019) into the productive university systems, such as the learning management system and the student management system.

Participants of the study were $N = 161$ students (106 female, 55 male) enrolled in a bachelor's program focusing on economic and business education at a European university. The average age of the participants was 22.97 years (SD = 2.92), and their average course load in the semester of data collection was six courses (SD = 2.11). Participants received 1 credit hour for participating in the study.

The course of study focused on quantitative research methods and included weekly face-to-face lectures as well as online materials through a learning management system such as lecture slides and video recordings, self-assessments, reading assignments, and discussion forums. In addition, students worked in small groups (max. 5) on a research project. The course concluded with a written exam at the end of the semester.

13.3.2 Data Collection Tools

Data used in this research consists of log data recorded through the universities LMS and learning analytics system LeAP. Data includes 9-week interactions of students with the digital learning environment. There are five different interaction themes: student-content, student-discussion, student-tutorial, student-video, and student-assessment. First, data preprocessing was applied. All data were cleaned while preprocessing and data quality checks were conducted. Then, the data were converted into contingency tables.

13.3.3 Data Analysis

Chi-square analysis is used when both the dependent and independent variables are discrete. However, in chi-square analysis, the number of categories in the variables is limited with two. Adjusted residuals analysis is used when the dependent and independent variables have more than two categories (Cornell Statistical Consulting Unit, 2018). Within the scope of this research, since there are contingency tables of 5 * 2, the standardized residuals analysis was used as the chi-square post hoc test. The standardized residual value should be $a < 0.05 \pm 1.96$ (Field, 2018). Bonferroni adjustment is made in order to test the significance level in these contingency tables. Bonferroni adjustment is carried out to maintain a sufficient error rate for contingency table dimensions (MacDonald & Gardner, 2000).

13.4 Results

The semester included 12 weeks of study; however, students' lived experience in the digital learning environment was limited to 9 weeks as through administrative constraints of the university. The learning experiences of students were examined as three different study periods. Each period is considered as 3 weeks (first period 1–3; second period 4–6; third period 7–9). The findings are presented based on these periods following the abovementioned research questions.

In the context of this research, students' interaction preference was examined not only gender but also achievement situation. Final course performances were used in order to determine the achievement situation of the students. Based on this performance, students were categorized as successful and unsuccessful. The threshold was handled 50% for this categorization.

13.4.1 First Period Interaction Preferences of Students

The first period includes the first 3 weeks (1–3 weeks) of students' digital learning environment interactions. The observed and expected interaction frequencies of the students in this period according to the gender and achievement are presented in Table 13.1.

As shown in Table 13.1, students interact intensely with the content. On a descriptive level, female students show higher interaction than male students, and similarly, successful students show higher interaction than unsuccessful students. In order to test for significant differences, adjusted residuals analysis has been

Table 13.1 Students' observed and expected interaction preferences in the first period

	Gender	Content	Discussion	Tutorial	Video	Assessment
Observed	Male	6558	1168	23	464	659
	Female	16,782	1142	7859	5291	710
Expected	Male	5093.28	504.09	1720.02	1255.86	298.74
	Female	18,246.72	1805.91	4499.14	4499.14	1070.26
	Achievement	**Content**	**Discussion**	**Tutorial**	**Video**	**Assessment**
Observed	Unsuccessful	6305	1146	25	701	493
	Successful	17,035	1164	7857	5054	876
Expected	Unsuccessful	4977.32	492.61	1680.86	1227.27	291.94
	Successful	18,362.68	1817.39	6201.14	4527.73	1077.06

Table 13.2 Results of adjusted residuals analysis about interaction preferences according to the gender for the first period

Gender	Content	Discussion	Tutorial	Video	Assessment
Male	**20.52**	**29.57**	−40.92	−22.34	**20.84**
Female	−10.84	−15.62	**21.62**	**11.81**	−11.01

Adjusted $p < 0.005$; z threshold >2.81

Table 13.3 Results of adjusted residuals analysis about interaction preferences according to the achievement for the first period

Achievement	Content	Discussion	Tutorial	Video	Assessment
Unsuccessful	**18.82**	**29.44**	−40.39	−15.02	**11.77**
Successful	−9.80	−15.33	**21.03**	**7.82**	−6.13

Adjusted $p < 0.005$; z threshold >2.81

Table 13.4 Students' observed and expected interaction preferences in the second period

	Gender	Content	Discussion	Tutorial	Video	Assessment
Observed	Male	52,025	976	47,940	3488	1435
	Female	66,439	5472	59,485	59,292	19,549
Expected	Male	39,674.26	2159.47	35,977.24	21,025.37	7027.66
	Female	78,789.74	4288.53	71,447.76	41,754.63	13,956.34
	Achievement	**Content**	**Discussion**	**Tutorial**	**Video**	**Assessment**
Observed	Unsuccessful	49,935	2735	52,674	15,572	2406
	Successful	68,529	3713	54,751	47,208	18,578
Expected	Unsuccessful	46,216.929	2515.589	41,910.231	24,492.663	8186.589
	Successful	72,247.071	3932.411	65,514.769	38,287.337	12,797.411

conducted. The findings obtained from this analysis are presented in Tables 13.2 and 13.3.

As shown in Table 13.2, the interaction preferences of students significantly differ based on their gender. The values which are presented in Table 13.2 are z scores. After Bonferroni adjusted in the analysis phase, the z threshold value was obtained as 2.81. Accordingly, it was determined that male students preferred to interact with content, discussion, and assessment in the first period. It can be said that female students prefer to interact with tutorials and videos compared to male students.

It can be said that students' interaction preferences significantly differ based on their achievement situation. According to the results, successful students preferred to interact with tutorial and video in the first period. On the other hand, unsuccessful students prefer to interact with content, discussion, and assessment.

13.4.2 *Second Period Interaction Preferences of Students*

The second period includes weeks 4–6 of the semester. The observed and expected interaction frequencies of the students' interactions in this period are presented in Table 13.4.

Table 13.5 Results of adjusted residuals analysis about interaction preferences according to the gender for the second period

Gender	Content	Discussion	Tutorial	Video	Assessment
Male	**62.01**	−25.47	**63.07**	−120.95	−66.71
Female	−44.00	**18.07**	−44.75	**85.82**	**47.34**

Adjusted $p < 0.005$; z threshold >2.81

Table 13.6 Results of adjusted residuals analysis about interaction preferences according to the achievement for the second period

Achievement	Content	Discussion	Tutorial	Video	Assessment
Unsuccessful	**17.29**	**4.37**	**52.58**	−57.00	−63.89
Successful	−13.83	−3.50	−42.05	**45.59**	**51.10**

Adjusted $p < 0.005$; z threshold >2.81

As shown in Table 13.4, students interact intensely with the content. On descriptive level, female students interact with tutorials and videos. Likewise, male students interact with not only content but also the tutorial. In addition, female students interact more frequently than male students. On the other hand, it's seen that successful students have higher interaction than unsuccessful students for all interaction types. Especially, successful students interact with video and assessment higher than unsuccessful students. In order to test for significant differences, adjusted residuals analysis has been conducted. The findings obtained from this analysis are presented in Tables 13.5 and 13.6.

As shown in Table 13.5, the interaction preferences of students significantly differ between female and male students. Accordingly, male students preferred to interact with content and tutorial in the second period. Further, female students prefer to interact with discussion, video, and assessments.

According to the results in Table 13.6, the interaction preferences of students significantly differ between successful and unsuccessful students. Successful students preferred to interact with video and assessment in the second period. Further, unsuccessful students prefer to interact with content, discussion, and tutorial.

13.4.3 Third Period Interaction Preferences of Students

The third period includes the last 3 weeks (7–9 weeks) of students' digital learning environment interactions. The observed and expected interaction frequencies of the students in this period are presented in Table 13.7.

As shown in Table 13.7, students interact intensely with the content. On a descriptive level, female students interact with tutorials and videos. In addition, female students interact more frequently than male students. While unsuccessful students interact with contents, successful students interact with tutorial and video. Adjusted residuals analysis was conducted to test for significant differences. The findings obtained from this analysis are presented in Tables 13.8 and 13.9.

Table 13.7 Students' observed and expected interaction preferences in the third period

	Gender	Content	Discussion	Tutorial	Video	Assessment
Observed	Male	33,311	291	18,106	6783	2014
	Female	80,178	5838	37,020	54,967	2574
Expected	Male	28,482.64	1538.21	13,835.12	15,497.56	1151.46
	Female	85,006.36	4590.79	41,290.88	46,252.44	3436.54
	Achievement	**Content**	**Discussion**	**Tutorial**	**Video**	**Assessment**
Observed	Unsuccessful	67,177	4287	24,068	27,715	2014
	Successful	46,312	1842	31,058	34,035	2574
Expected	Unsuccessful	58,966.43	3184.49	28,642.279	32,083.96	2383.82
	Successful	54,522.56	2944.50	26,483.72	29,666.03	2204.17

Table 13.8 Results of adjusted residuals analysis about interaction preferences according to the gender for the third period

Gender	Content	Discussion	Tutorial	Video	Assessment
Male	**28.61**	−31.80	**36.31**	−70.00	**25.42**
Female	−16.56	**18.41**	−21.02	**40.52**	−14.71

Adjusted $p < 0.005$; z threshold >2.81

Table 13.9 Results of adjusted residuals analysis about interaction preferences according to the achievement for the third period

Achievement	Content	Discussion	Tutorial	Video	Assessment
Unsuccessful	**40.99**	**22.03**	**12.53**	−50.95	−62.02
Successful	−35.62	−19.15	−10.89	**44.27**	**53.90**

Adjusted $p < 0.005$; z threshold >2.81

As shown in Table 13.8, the interaction preferences of students significantly differ based on their gender. Accordingly, male students preferred to interact with content, tutorial, and assessment in the third period, and female students prefer to interact with discussion and video.

When we examined the results, it's possible to say the interaction preferences of students significantly differed based on their achievement situation. According to the results, successful students preferred to interact with video and assessment in the third period. On the other hand, unsuccessful students prefer to interact with content, discussion, and tutorial.

After the students' interaction preferences for three different periods are given based on gender and achievement, the interaction preferences for each gender are presented in Table 13.10, and the interaction preferences for achievement situation are presented in Table 13.11.

As shown in Table 13.10, the interaction preferences of students differ throughout the study periods according to their gender. For example, male students preferred to interact with content, discussion, and assessment in the first period. However, in the second period, male students chose to interact with content and tutorial. According to these findings, it is possible to say that students have different

Table 13.10 Student interaction preferences according to their gender

Gender	First period	Second period	Third period
Male	Content Discussion Assessment	Content Tutorial	Content Tutorial Assessment
Female	Tutorial Video	Discussion Video Assessment	Discussion Video

Table 13.11 Student interaction preferences according to achievement

Achievement	First period	Second period	Third period
Unsuccessful	Content Discussion Assessment	Content Discussion Tutorial	Content Discussion
Successful	Tutorial Video	Video Assessment	Tutorial Video Assessment

interaction preferences in different periods. Another point that draws attention to the findings is that female students consistently prefer to interact with video and male students to interact with content.

As shown in Table 13.11, the interaction preferences of students differ throughout the study periods according to their achievement. For example, successful students preferred to interact with tutorial and video in the first period. However, in the second period, these students preferred to interact with video and assessment. When we look at the third period, successful students preferred to interact with tutorial, video, and assessment. According to these findings, it is possible to say that students have different interaction preferences in different periods. Another point that draws attention to the findings is that successful students consistently prefer to interact with video and unsuccessful students to interact with content and discussion.

13.4.4 All Period Interaction Preferences of Students

In addition to these results, students' interactions for 9 weeks were also examined as a whole in order to provide a general perspective. The observed and expected interaction frequencies of the students during the whole learning experience are presented in Table 13.12.

As shown in Table 13.12, students interact intensely with the content and tutorial. On a descriptive level, female students interact with content, tutorials, and videos. In addition, female students interact more frequently than male students. When we look at the interaction preferences according to the students' achievement situation, successful students interact with video and assessment. Besides this, it's also seen that successful students interact more frequently than unsuccessful students.

Table 13.12 Students' observed and expected interaction preferences in the all periods

	Gender	Content	Discussion	Tutorial	Video	Assessment
Observed	Male	91,894	2435	66,069	10,735	4108
	Female	163,399	12,452	104,364	119,550	22,833
Expected	Male	74,832.52	4363.74	49,958.01	38,189.67	7897.06
	Female	180,460.48	10,523.26	120,474.99	92,095.33	19,043.94
	Achievement	**Content**	**Discussion**	**Tutorial**	**Video**	**Assessment**
Observed	Unsuccessful	123,417	8168	76,677	43,988	4913
	Successful	131,876	6719	93,666	86,297	22,028
Expected	Unsuccessful	109,831.91	6404.67	73,284.80	56,051.09	11,590.53
	Successful	145,461.09	8482.33	97,058.20	74,233.91	15,350.47

Table 13.13 Results of adjusted residuals analysis about interaction preferences according to the gender for the all periods

Achievement	Content	Discussion	Tutorial	Video	Assessment
Male	**62.37**	−29.20	**72.08**	−140.49	−42.64
Female	−40.16	**18.80**	−46.42	**90.47**	**27.46**

Adjusted $p < 0.005$; z threshold >2.81

Table 13.14 Results of adjusted residuals analysis about interaction preferences according to the achievement for the all periods

Achievement	Content	Discussion	Tutorial	Video	Assessment
Unsuccessful	**40.99**	**22.03**	**12.53**	−50.95	−62.02
Successful	−35.62	−19.15	−10.89	**44.27**	**53.90**

Adjusted $p < 0.005$; z threshold >2.81

Adjusted residuals analysis was applied to test for significant differences. The findings obtained from this analysis are presented in Tables 13.13 and 13.14.

As shown in Table 13.13, the interaction preferences of students significantly differ based on their gender. Accordingly, male students preferred to interact with content and tutorial; female students prefer to interact with discussion, video, and assessment.

When we examined the results in Table 13.14, it is possible to confirm that the interaction preferences of students significantly differed based on their achievement situation. According to the results, successful students preferred to interact with video and assessment. On the other hand, unsuccessful students prefer to interact with content, discussion, and tutorial.

13.5 Discussion and Conclusion

In this study, interaction preferences of students in a digital learning environment were examined based on their gender. For this purpose, the standardized residuals analysis was conducted as the chi-square post hoc test. According to the findings, it was concluded that students' online interaction preferences are different based on their gender, achievement situation (successful/unsuccessful), and study period. The interaction preferences of male students are as follows: (a) for the first period, content, assessment, and discussion; (b) for the second period, content and tutorial; and (c) for the third period, content, tutorial, and assessment. The interaction preferences of female students are as follows: (a) for the first period, tutorial and video; b) for the second period, discussion, video, and assessment; and (c) for the third period, discussion and video. The interaction preferences of successful student are as follows: (a) for the first period, tutorial and video; (b) for the second period, video and assessment; and (c) for the third period, video and assessment. The interaction preferences of unsuccessful student are as follows: (a) for the first period, content, discussion, and assessment; (b) for the second period, content, discussion, and tutorial; and (c) for the third period, content, discussion, and tutorial.

Another finding obtained within the scope of the research points toward the preference of male students being the interaction with the textual content, and female students prefer to interact with the video content. This finding is consistent with Ching and Hsu (2015) who found that female students prefer to interact with audio/video discussions and male students prefer text-based discussions. Female students seem to prefer to interact with video content because of including more personal and emotional forms of language (Ching & Hsu, 2015). However, more studies are needed to foster these findings on large scale. On the other hand, female students tend to interact with discussion more than male students. As the next step, the source of this behavior can be investigated. In this context, it is recommended to investigate students' social anxiety levels in digital learning environments. Social anxiety affects the students' interaction with the digital learning environments (Agarwal & Karahanna, 2000).

In addition to these findings, successful students prefer to interact with video, and unsuccessful students prefer to interact with content and discussion. In the digital learning environments hypothetically, it is expected that students, respectively, interact with content to obtain the knowledge, interact with the discussion to construct the knowledge, and interact with the assessment to reflect the knowledge (Şahin, Keskin, & Yurdugül, 2020). When we examined the results, it is possible to say that unsuccessful students have a difficulty with constructing knowledge. Therefore, they interact extensively with the discussion in order to structure knowledge. In this context, it is suggested that some guidance, recommendations, and interventions should be integrated to the digital learning environments. This may be realized through prompts using learning analytics approaches (Schumacher & Ifenthaler, 2021). Besides this, successful students tend to reflect the knowledge via interacting with assessment. For the further research, students' help-seeking behavior should be investigated. Help-seeking behavior affects the students' learning process (Arroyo et al., 2004) and has a positive effect in the context of the digital learning environments (Aleven et al., 2003). In addition to these, successful students tend to interact with video, and unsuccessful students tend to interact with the content. In other words, it can be said that students who interact with video are more successful. Video affects students' learning performance positively (Zhang et al., 2006). However, more studies and experimental evidence are needed to foster these findings.

Findings obtained in this research can provide some initial suggestions for instructional designers of digital learning environments (Ifenthaler, 2017). Specifically, personalized learning paths and recommendations for interacting with other learning materials available may be easy to implement features of future digital learning environments. The findings of this study are limited because the interaction preferences of students have been reduced to gender and performance. In further research, it is planned that students' individual dispositions will be examined, such as learning strategies, cognitive strategies, social anxiety, help-seeking behavior, motivational dispositions, and prior knowledge. In this way, it is thought that further recommendations can be obtained for researchers and learning designers in order to design appropriate digital learning environments.

References

Agarwal, R., & Karahanna, E. (2000). Time flies when you're having fun: Cognitive absorption and beliefs about information technology usage. *MIS Quarterly, 24*(4), 665–694. https://doi.org/10.2307/3250951

Aleven, V., Stahl, E., Schworm, S., Fischer, F., & Wallace, R. (2003). Help seeking and help design in interactive learning environments. *Review of Educational Research, 73*(3), 277–320.

Arroyo, I., Murray, T., Woolf, B. P., & Beal, C. (2004). Inferring unobservable learning variables from students' help seeking behavior. In *International Conference on Intelligent Tutoring Systems* (pp. 782–784). Springer.

Behrens, J., Mislevy, R., Dicerbo, K., & Levy, R. (2012). Evidence centered design for learning and assessment in the digital world. In M. Mayrath, J. Clarke-Midura, D. Robinson, & G. Schraw (Eds.), *Technology-based assessments for 21st century skills* (pp. 13–54). Information Age Publishers.

Blagojević, M., & Milosević, M. (2013). Collaboration and learning styles in pure online courses: An action research. *Journal of Universal Computer Science, 19*(7), 984–1002.

Brown, M., & Long, P. (2006). Trends in learning space design. *Learning Spaces, 9*, 1–9.

Brown, M., Dehoney, J., & Millichap, N. (2015). The next generation digital learning environment. *A Report on Research. ELI Paper.* Educause.

Brusilovsky, P., & Henze, N. (2007). Open corpus adaptive educational hypermedia. In P. Brusilovsky, A. Kobsa, & W. Nejdl (Eds.), *The adaptive web: Methods and strategies of web personalization* (pp. 671–696). Springer.

Ching, Y. H., & Hsu, Y. C. (2015). Online graduate students' preferences of discussion modality: Does gender matter? *Journal of Online Learning and Teaching, 11*(1), 31–41.

Cornell Statistical Consulting Unit. (2018). *Adjusted standardized residuals for interpreting contingency tables* (Report No. 95). Retrieved from https://www.cscu.cornell.edu/news/statnews/stnews95.pdf

Field, A. (2018). *Discovering statistics using IBM SPSS statistics*. SAGE Publications.

Graf, S., Kinshuk, & Liu, T. C. (2009). Supporting teachers in identifying students' learning styles in learning management systems: An automatic student modelling approach. *Journal of Educational Technology and Society, 12*(4), 3–14.

Ifenthaler, D. (2012). Learning management systems. In N. M. Seel (Ed.), *Encyclopedia of the sciences of learning* (Vol. 12, pp. 1925–1927). Springer.

Ifenthaler, D. (2015). Learning analytics. In J. M. Spector (Ed.), *The SAGE encyclopedia of educational technology* (Vol. 2, pp. 447–451). Sage.

Ifenthaler, D. (2017). Learning analytics design. In L. Lin & J. M. Spector (Eds.), *The sciences of learning and instructional design. Constructive articulation between communities* (pp. 202–211). Routledge.

Ifenthaler, D., & Pirnay-Dummer, P. (2011). States and processes of learning communities. Engaging students in meaningful reflection and learning. In *Social media tools and platforms in learning environments* (pp. 81–94). Springer.

Ifenthaler, D., & Widanapathirana, C. (2014). Development and validation of a learning analytics framework: Two case studies using support vector machines. *Technology, Knowledge and Learning, 19*(1–2), 221–240.

Ifenthaler, D., & Yau, J. Y.-K. (2020). Utilising learning analytics to support study success in higher education: A systematic review. *Educational Technology Research and Development, 68*(4), 1961–1990. https://doi.org/10.1007/s11423-020-09788-z

Ifenthaler, D., Pirnay-Dummer, P., & Seel, N. M. (Eds.). (2010). *Computer-based diagnostics and systematic analysis of knowledge*. Springer.

Ji, H., Park, K., Jo, J., & Lim, H. (2016). Mining students activities from a computer supported collaborative learning system based on peer to peer network. *Peer-to-Peer Networking and Applications, 9*(3), 465–476.

Jo, I. H., Yu, T., Lee, H., & Kim, Y. (2015). Relations between student online learning behaviour and academic achievement in higher education: A learning analytic approach. In G. Chen, V. Kumar, Kinshuk, R. Huang, & S. C. Kong (Eds.), *Emerging issues in smart learning* (pp. 275–286). Springer.

Keskin, S., & Yurdugül, H. (2019). Factors affecting students' preferences for online and blended learning: Motivational vs. cognitive. *European Journal of Open, Distance and E-learning, 22*(2), 72–86.

Keskin, S., Şahin, M., & Yurdugül, H. (2019). Online learners' navigational patterns based on data mining in terms of learning achievement. In *Learning technologies for transforming large-scale teaching, learning, and assessment* (pp. 105–121). Springer.

Klasen, D., & Ifenthaler, D. (2019). Implementing learning analytics into existing higher education legacy systems. In D. Ifenthaler, J. Y.-K. Yau, & D.-K. Mah (Eds.), *Utilizing learning analytics to support study success* (pp. 61–72). Springer.

Land, S., & Jonassen, D. H. (2009). Student-centered learning environments. In D.H. Jonassen & S. Land (Eds.), *Theoretical foundations of learning environments*. Routledge.

Land, S. M., & Hannafin, M. J. (1997). Patterns of understanding with open-ended learning environments: A qualitative study. *Educational Technology Research and Development, 45*(2), 47–73.

MacDonald, P. L., & Gardner, R. C. (2000). Type I error rate comparisons of post hoc procedures for I j Chi-Square tables. *Educational and Psychological Measurement, 60*(5), 735–754.

Macfadyen, L. P., & Dawson, S. (2010). Mining LMS data to develop an—Early warning system for educators: A proof of concept. *Computers and Education, 54*, 588–599.

Machado, M., & Tao, E. (2007). Blackboard vs. Moodle: Comparing user experience of learning management systems. In *2007 37th annual frontiers in education conference-global engineering: Knowledge without borders, opportunities without passports* (pp. S4J-7). IEEE.

Moore, M. G. (1989). Three types of interaction. *The American Journal of Distance Education, 3*(2), 1–7.

Northrup, P. T. (2009). Online learners' preferences for interaction. In *The perfect online course: Best practices for designing and teaching* (pp. 463–473).

Rhode, J. (2009). Interaction equivalency in self-paced online learning environments: An exploration of learner preferences. *The International Review of Research in Open and Distributed Learning, 10*(1), 1–23.

Şahin, M., Keskin, S., & Yurdugül, H. (2020). Sequential analysis of online learning behaviors according to E-learning readiness. In P. Isaias, D. G. Sampson, & D. Ifenthaler (Eds.), *Online teaching and learning in higher education* (pp. 117–131). Springer.

Şahin, M., Keskin, S., Özgür, A., & Yurdugül, H. (2017). Determination of interaction profiles based on learner characteristics in e-learning environment. *Educational Technology Theory and Practice, 7*(2), 172–192.

Schumacher, C., & Ifenthaler, D. (2018). The importance of students' motivational dispositions for designing learning analytics. *Journal of Computing in Higher Education, 30*(3), 599–619. https://doi.org/10.1007/s12528-018-9188-y

Schumacher, C., & Ifenthaler, D. (2021). Investigating prompts for supporting students' self-regulation—A remaining challenge for learning analytics approaches? *The Internet and Higher Education, 49*, 100791. https://doi.org/10.1016/j.iheduc.2020.100791

Schumacher, C., Klasen, D., & Ifenthaler, D. (2019). Implementation of a learning analytics system in a productive higher education environment. In M. S. Khine (Ed.), *Emerging trends in learning analytics* (pp. 177–199). Brill.

Šimić, G., Gašević, D., & Devedžić, V. (2004). Semantic web and intelligent learning management systems. In *Workshop on Applications of Semantic Web Technologies for e-Learning*.

Zhang, D., Zhou, L., Briggs, R. O., & Nunamaker, J. F., Jr. (2006). Instructional video in e-learning: Assessing the impact of interactive video on learning effectiveness. *Information & Management, 43*(1), 15–27.

Chapter 14
Development of Fill-in Workbook System to Visualize Learning Attitude

Kousuke Abe, Tetsuo Tanaka, and Kazunori Matsumoto

14.1 Introduction

The computerization of higher education is progressing with, for example, the spread of LMSs (learning management systems), which manage lecture materials and student reports on the Internet. In conjunction with this, learning analytics, which analyzes operation logs in LMSs and e-learning systems, is attracting attention. Learning analytics involves collecting a log of the learning process, to identify items such as which part of the material the student is learning and when and for how long the student is learning. Analyzing this can contribute to class improvement and the provision of learning support (Ferguson, 2012; Kumar & Vivekanandan, 2018; Viberg et al., 2018; Diana et al., 2019; Verbert et al., 2020). Particularly in recent years, a real-time learning analysis environment that supports teachers and students in classrooms has been made possible by the development of wireless LANs in classrooms and the practice of "bring your own device" (BYOD), in which students bring their own laptops and other devices into the classroom (Shimada et al., 2018; Shimada & Konomi, 2017; Fu et al., 2017).

The operation logs used in learning analytics mainly contain data on time and content such as turning the pages of teaching materials, underlines, memos, and questions. Therefore, it is possible to ascertain the status of engaged learners who actively underline, take notes, or ask questions. However, it is difficult to ascertain the status of many other students who attend classes but do not participate actively in learning.

Based on this background, the authors are developing a fill-in workbook system that will both allow faculty to ascertain the learning attitudes of all students,

K. Abe · T. Tanaka (✉) · K. Matsumoto
Kanagawa Institute of Technology, Atsugi, Japan
e-mail: t-tanaka@ic.kanagawa-it.ac.jp

D. Ifenthaler et al. (eds.), *Orchestration of Learning Environments in the Digital World*,
Cognition and Exploratory Learning in the Digital Age,
https://doi.org/10.1007/978-3-030-90944-4_14

including those who are not active, and improve lectures through well-timed and appropriate actions and also help students to understand lectures in an efficient manner. In a class using this system, students are provided with materials containing blanks to be filled in, and they engage in the lesson while filling the blanks in the materials. The operation logs are collected and analyzed and teachers and students receive feedback. This system compels students in class to interact with the system and makes it possible to collect the learning activities of students who are not active in class.

The authors have already reported on the function and configuration of the system and the results of a trial using the system in a small class. A trial with a class of about 100 students and the subsequent extraction of class attitudes from system logs were reported in Abe et al. (2018) and Abe et al. (2019). In Abe et al. (2020a), we reported the results of using this system continuously over multiple lessons of a single course. In Abe et al. (2020b), we reported on prototyping a retrospective review function of this system.

This paper describes the functions and configurations of the improved system based on the results of the trials so far and its evaluation. In Sect. 14.2, we summarize the targeted lecture and associated problems. Section 14.3 describes an approach to solve the problems. Section 14.4 describes the functions of the system, and Sect. 14.5 describes the implementation. The results used in the actual lesson are described in Sect. 14.6.

14.2 Targeted Lecture and Associated Problems

In this study, we focus on face-to-face lecture classes that are commonly held at private Japanese universities. Classes are generally held in classrooms that accommodate about 100 students, and these classrooms are often held in a classroom with raked seating as shown in Fig. 14.1. The teacher stands at the front on a stage and lectures while writing on a board or using slides prepared with presentation software. Students listen to lectures and write down essential facts in their textbooks and take notes.

If the time spent listening to the teacher's explanation becomes longer in such a large class, the number of students who cannot concentrate on the teacher's explanation will increase.

In order to sustain the concentration of students, faculty members give lessons using the so-called Monta method (Reynolds, 2005), which is a presentation method in which important words and phrases are left blank as shown in Fig. 14.2 and the blanks are filled in for each explanation. By writing (filling in the blanks) the blank contents displayed for each explanation in the pre-distributed paper materials, the students will be constantly moving their hands while they take the course, and their concentration will continue for longer. When listening to students' opinions about the lessons using the Monta method in the minute report (Schwartz, 1991),

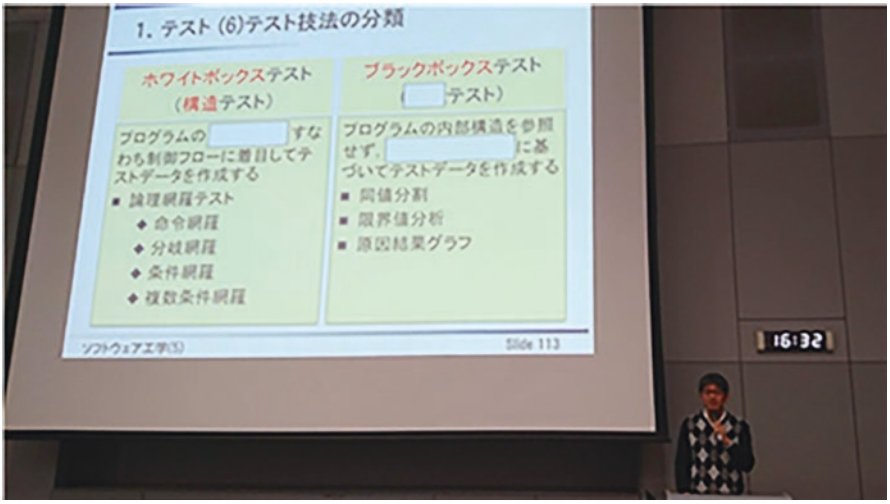


Fig. 14.2 Lecture using the Monta method

Fig. 14.1 Lecture in a classroom with raked seating

comments were heard such as, "I am always moving my hands so I don't get sleepy" and "I keep concentrating."

However, in a large number of classes, it is difficult to ascertain how many students are engaged in filling in the blanks. Especially in a classroom with raked seating, teachers do not know how many students are taking notes because the students' hands are usually not visible. The lecturer will move on to the next step while trying to gauge whether the students are keeping up, but if the lecture proceeds too quickly, the following requests for improvement are made by the students: "The lecture moves too quickly on to the next step" and "I cannot digest what is being said if it is spoken while I am still filling in the blanks."

Many teachers provide exercises during class to give students time to think in order to maintain their concentration. One of the authors asks the students to solve fill-in-the-blank exercises, and when they have finished, the author displays the correct answers and lets the students check how they did. There are also student comments about this, such as "I can check my level of understanding" and "I got a clear idea after doing the exercises." Such exercises are useful in helping students to understand the lesson.

However, in a large class, it is not possible to ascertain how many students are working on the exercise or how many students are progressing to what problem, so it is difficult for teachers to take actions, such as giving hints when progress is poor. In addition, in this exercise method, students are divided between those who actively engage in the exercise and those who copy the answers after the teacher displays them. However, it is not possible to distinguish between students who have finished answering and students who have given up and are not engaged in the exercise. The teachers move on to answering while watching the students, but if it is too soon, the students who are still actively working on it say that "the time for thinking in the exercises is too short," and if a teacher does not allow enough time after displaying the answers, students who are copying them said, "I missed some of the answers to the exercise."

14.3 Approach

In this study, in order to record the learning behavior of students, instead of distributing paper materials, a fill-in-the-blank workbook will be published on the Web. Students refer to the workbook via a PC or tablet Web browser, listen to lectures, and solve exercises while filling in the blanks. In addition, the underline and memo writing that were done in paper handouts are also done in the browser. In addition, students can enter questions and comments that were previously done on minute papers from the same screen.

The teaching materials are PDF documents that teachers are accustomed to using. Our approach provides an environment for hassle-free conversion of PDF documents into fill-in-the-blank workbooks and collects learning behavior logs for activities such as fill-in-the-blanks, underlines, and memos.

14.3.1 Converting PDF Documents to Fill-in-the-Blank Workbooks

When the teacher creates a PDF document and registers it in the system, the student can display the PDF document in a Web browser and write notes such as questions, opinions, and impressions on the PDF document or underline certain sections. In addition, teachers can use a Web browser to set blanks in the document. This allows teachers to create teaching materials using the familiar Word and PowerPoint applications. In addition, existing teaching materials can be used as they are as a fill-in-the-blank workbook. Independently of the PDF document, blanks/notes to the document/underlines can be displayed/added/edited/deleted, so the document is displayed in a four-layer structure as shown in Fig. 14.3.

14.3.2 Collection of Learning Behavior Logs

When the system is in use, students are compelled to interact with it, and the learning behavior logs of inactive students (students who do not actively participate in learning activities by, for example, underlining and writing notes) can be acquired. If teachers can effectively assess the students' learning attitude in real time during class, they can then take actions that encourage students to concentrate whenever the number of students with a poor learning attitude is seen to increase. Table 14.1 shows the items in the learning behavior logs collected using this system.

Fig. 14.3 Four-layer structured workbook

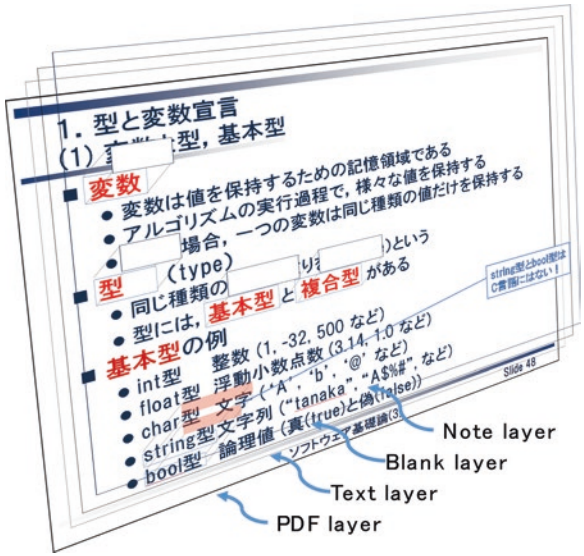


Table 14.1 The items in the learning behavior logs

Item	Description
Fill-in blanks	Content and time of each input to the blank space and whether the input is correct or incorrect
Underlining	Underline position and time
Notes	Note position, content, and time
Operation logs	Page focus time per minute, number of mouse operations per minute, number of keyboard operations per minute, start and end times of the period
Page turning	Destination page number and time
Question	Content of questions asked by students during class and time

Table 14.2 The items in the teacher's operation log

Item	Description
Page turning	Content and time of each answer written in the blank space and whether answers are correct or incorrect
Blanks revealed	Reveal (displaying the correct answer) times and identifiers of blanks revealed

In addition, a log of the teacher's operations is also compiled and can be used to compare course progress with the student's operation log. Table 14.2 shows the teacher's operation log.

14.3.3 Extraction of Learning Attitude

In order to enable teachers to adjust the lesson speed, the percentage of students who have completed filling in a section is calculated from the fill-in-the-blank log for each blank and presented to the teacher in real time as the "fill-in-the-blank rate." This allows teachers to adjust the timing of their move to the next page. In addition, the timing of answer matching can be adjusted during the exercise. Also, when the progress of the exercise is poor, hints can be given while the students are working on it.

In order to enable teachers to take action according to the trends of the entire classroom, changes in the attitude of each student during the class are extracted. We focus on unfocused time, page transitions, number of keyboard operations, and number of mouse clicks to extract the attitude of each student during the class as follows:

- A student who is synchronized with the teacher's page transition most of the time and for whom there is no unfocused time logged is considered to be attentive. Unfocused time is the time during which students do not focus on the system, for example, when they are opening another application such as to check email.

Table 14.3 Correspondence between operation log and learning attitude

	Learning attitude	Unfocused time	Keystroke, Mouse click	Page transition
1. Excellent	Following and concentrated	Less than 30 s per minute	–	Synchronized
2. Good	Following but not concentrated	30 s or moreper minute	–	Synchronized
3. Fair	Jumping ahead	–	–	Jumping ahead
4. Poor	Falling behind	–	Yes	Falling behind
5. Bad	Not participating	–	No	Falling behind

- The reference page may be synchronized with the teacher's explanation even in a time period with many unfocused periods within it. In addition, keyboard operations are also being performed a short time after page transition. The student at this time seems to be in a state of looking at the material without listening to the teacher's explanation and filling in the blanks on the page while looking at the teacher's presentation material immediately after turning the page.
- Students jumping ahead are not listening to the teacher's explanation, that is, their learning attitude can be regarded as bad.
- Students falling behind can also be considered as not being able to keep up with the pace of the lecture or sleeping or doing other things and not listening to the teacher's explanation. These can be distinguished by the presence or absence of operations.
- There may be a time period during which there is no page transition or operation for an extended period. During this time, the pages being explained by the teacher contain blanks, so it is necessary to fill in the gaps, but there are no operations and no page transition. It can be assumed that during this time the student is doing other things or sleeping without paying attention to the lecture.

To summarize the above, students' learning attitudes can be divided into the five categories shown in Table 14.3 based on unfocused time, page transitions, keyboard operations, and mouse clicks.

This allows teachers to adjust the pace of their lessons according to whether more students are ahead or behind. In addition, when the number of students who are not participating increases, it is possible to take actions to encourage the concentration of students, such as having them discuss with each other and imposing exercises. Furthermore, after class, the teacher can check the transition of the learning attitude of the entire class and identify points for improvement. Students can also look back on their class attitude.

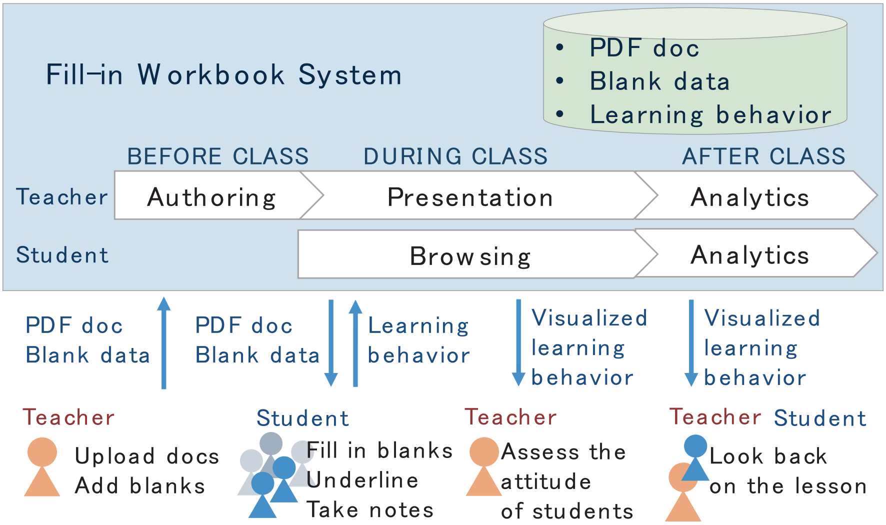


Fig. 14.4 System overview

14.4 Function of the System

In this research, based on the approach described in Sect. 14.3, we develop a lecture support system using a fill-in textbook. As shown in Fig. 14.4, this system provides four modes, an authoring mode in which teachers upload a PDF document and create teaching materials by adding blank spaces, a browsing mode in which students refer to teaching materials while filling in the blanks, a presentation mode in which teachers explain teaching materials while referring to the student's learning attitude, and an analytics mode to support a retrospective review of teachers and students. Details of each mode will be described below. Moreover, it has user management function and digital textbook management function like other lecture support systems.

14.4.1 Authoring Mode

The authoring mode is a mode in which teachers upload a PDF document and create teaching materials by adding blank spaces. The teachers set a blank by specifying the character string that he/she wants to blank and clicking the blank button. In order to automatically match the answers when the student fills in the blanks, the blank identifier and specified character string are registered automatically in the system in advance.

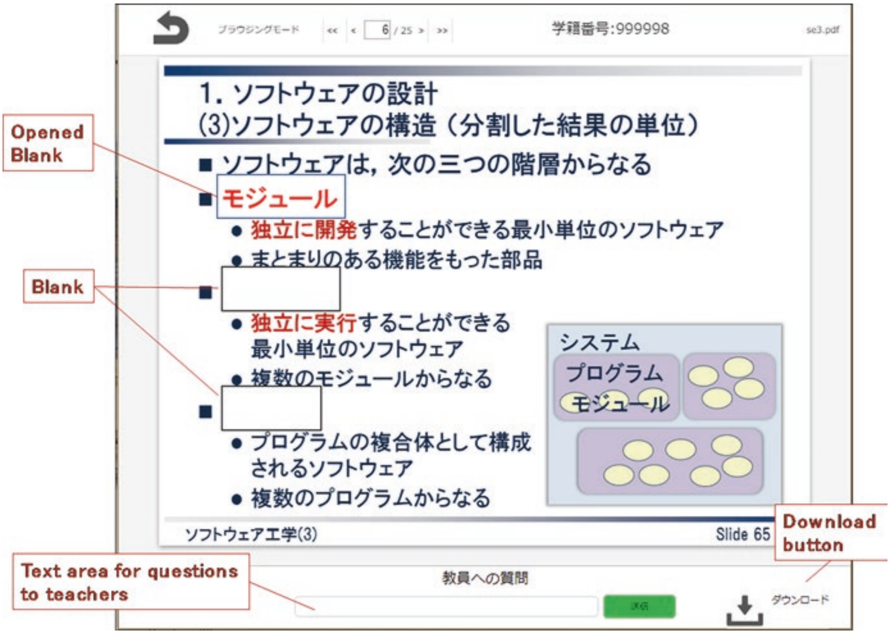


Fig. 14.5 Example of the student's screen

This allows the student to immediately know whether his/her input is correct or not. In addition, teachers do not need to register the correct answer corresponding to the blank.

14.4.2 Browsing Mode

The browsing mode is a mode in which students refer to teaching materials while filling in the blanks during classes. Students can take notes, fill in blanks, underline, and ask the teacher questions while browsing PDF documents on the screen as shown in Fig. 14.5. Besides the display pane of the PDF document, there are a comment pane for the page and a controller pane with buttons for turning pages, underlining, and comment for the specific words.

In the browsing mode, operations such as page turns and blank completions, underlines, comments, keyboard presses, mouse clicks, focus and blur to the window of the system, and the time thereof are accumulated in the system as a log of learning behavior.

Fig. 14.6 Example of the teacher's screen

14.4.3 Presentation Mode

The presentation mode is a mode in which teachers explain teaching materials while referring to the student's learning attitude visualized in real time on the presentation mode screen.

This mode consists of two screens: a presentation screen displaying only the slide and a learning behavior visualization screen displaying slides and various data as shown in Fig. 14.6. It is designed to show the presentation screen to the students by projector and to display the learning behavior visualization screen on the laptop PC of the teacher giving the lecture.

If the teacher clicks on one of the two screens, the blank space disappears according to the deletion order set in the authoring mode. If there are no blank spaces to be deleted in the slide, the screen shifts to the next slide. Also, the two screens transit synchronously, i.e., when blank deletion or slide transition is performed on one screen, the other transitions as well.

In the learning behavior visualization screen of Fig. 14.2, student learning behavior logs are visualized in real time. The number of students interacting with the system, the students' progress in filling in blanks, the heat-up degree, the question contents, the description contents to the note, and the underlined character strings are sent to the teacher's screen in real time. By referring to them, the teacher can lecture in a way that takes into account the learning behavior of the students, such as by adjusting the speed of the lecture according to the students' progress in filling in blanks or by giving hints appropriately in response to problems such as a poor correct answer rate during an exercise.

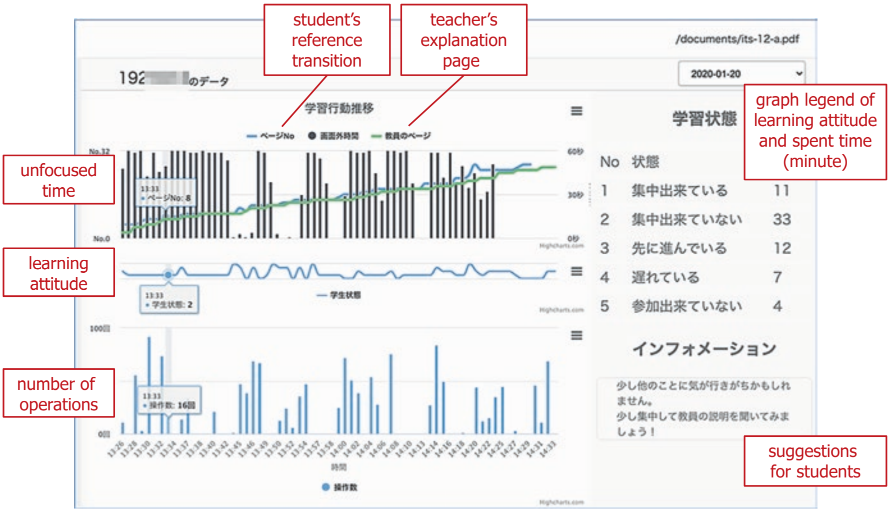


Fig. 14.7 Example of the retrospective review screen for students

14.4.4 Analytics Mode for Students

By presenting the transition of learning attitude to the students in the form of a retrospective review after the lecture, they can be made aware of the problems detected during the lecture and become more conscious of their learning attitude. The analytics mode is the mode for achieving this.

Figure 14.7 shows an example of the retrospective review screen. The horizontal axis is the time from the start of the lesson, measured in minutes. The bar graph at the top shows the unfocused time; the green line on the graph indicates the page being explained by the teacher, and the gray line on the graph indicates the page being referenced by the student. The left vertical axis corresponds to the line graph showing page numbers, and the right vertical axis corresponds to the bar graph showing unfocused time (seconds). The middle part of Fig. 14.7 is a line graph showing learning attitude. The bottom bar graph shows the total number of mouse clicks and keyboard operations per minute.

This allows students to look back on their class attitude after class. For example, the students can find out in which part of the lesson they could not keep up with the teacher's explanation or if/when they had a lapse in concentration. Thereby they can focus on reviewing that part or ask other students/teachers about that part.

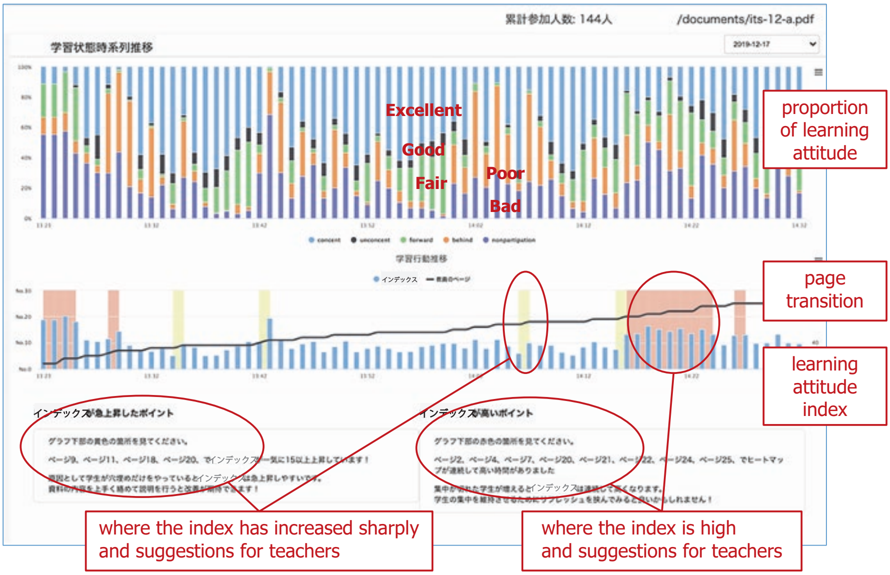


Fig. 14.8 Example of the retrospective review screen for teachers

14.5 Analytics Mode for Students

This system provides the function of creating a time sequence aggregate learning attitude chart by totaling the learning attitudes of all students described in the previous section and the function of presenting it to the teacher. This will allow teachers to improve their lessons. For example, if many students are falling behind in the class, the teacher can slow down the class or explain the material in more detail. After class, the teacher can improve teaching materials that are in use when many students are falling behind. In addition, individual guidance can be given such as selecting students who tend to fall behind and giving them advice.

Figure 14.8 shows a review screen for teachers. The horizontal axis is time. The upper bar graph shows the ratio of the five learning states at each moment in time. This allows the teachers to understand how the student's learning status is changing. The lower bar graph shows the value of the learning attitude index (LAindex) expressed by the following formula at each moment in time:

$$\text{LAindex} = \frac{\left(\text{LA}_2 + 2\text{LA}_3 + 3\text{LA}_4 + 4\text{LA}_5\right)}{4N} \times 100$$

LA_i: number of students of status i in Table 14.1
N: number of students

Furthermore, the points where LAindex sharply rises are displayed in yellow, and the points where LAindex continues at a high level are displayed in pink. These are the points that teachers should pay attention to.

14.6 Implementation

This system was developed as a Web application. Figure 14.9 shows its configuration. The client-side and server-side implementations are described below.

Node.js, which has abundant useful packages, is used to implement the server, and Express is used as the framework. RDBMS MySQL is used to save and manage student learning behavior logs, account information, and fill-in-the-blank workbooks. For client-server communication, HTTP communication, Socket.io for two-way real-time communication, and Super Agent for the purpose of simplifying Ajax communication for asynchronous communication are used. Socket.io is used for communication that requires real-time transmission and reception, that is, transmission of the learning behavior log collected in the browsing mode to the server and transmission of the learning attitude in the presentation mode from the server to the teacher client. Super Agent is used for communication that does not emphasize real-time performance, such as setting blanks in the authoring mode and registering subject information. In addition, HTTP communication is used for page transitions.

HTML5, CSS, and JavaScript are used for client-side implementation. Also, PDF.js is used to display PDF on the browser. This makes it possible to select a character string in the PDF.

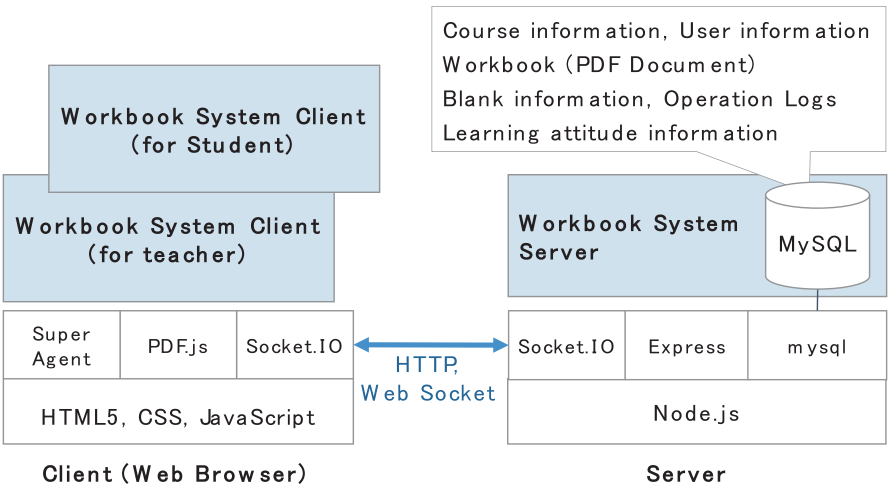


Fig. 14.9 System configuration

14.7 Evaluation of the System

The authors experimentally evaluated whether or not teachers could ascertain the learning attitudes of students using this system. Specifically, we conducted actual classes using this system for students in the Department of Computer Science at our university and conducted interview surveys with teachers and questionnaire surveys with students. The target classes were in the following three subjects:

- Software Engineering (2019, 2020)
- Approximately 90 second-year students in the Department of Computer Science.
- The eighth class in 2020 was held with the fill-in-the-blank rate hidden.
- Introduction to Software (2019, 2020)
- Approximately 130 first-year students in the Department of Computer Science.
- The 13th class in 2020 was held with the time-series heat map hidden.
- Operating System (2020)
- Approximately 90 second-year students in the Department of Computer Science.
- The 13th class in 2020 was held with the time-series heat map hidden.

14.7.1 Teacher's Explanation Time

To investigate whether the visualization of learning attitudes by this system affects the explanation time of teachers, we compared the explanation time of the teacher when the fill-in-the-blank rate and time-series heat map were displayed and when they were not displayed in the same class for the same subject. We used the data from the classes of two subjects (software engineering and software basic theory) that had been using the system since 2019. The display/non-display of the fill-in-the-blank rate and time-series heat map were as follows:

1. 2019: Show fill-in-the-blank rate and time-series heat maps for all classes.
2. 2020:
   (a) Software Engineering Seventh Class: Show fill-in-the-blank rate and time-series heat map.
   (b) Software Engineering eighth Class: Hide fill-in-the-blank rate; display time-series heat map.
   (c) Software Basics 11th Class: Show fill-in-the-blank rate and time-series heat map.
   (d) Software Basics 13th Class: Show fill-in-the-blank rate; hide time-series heat map.

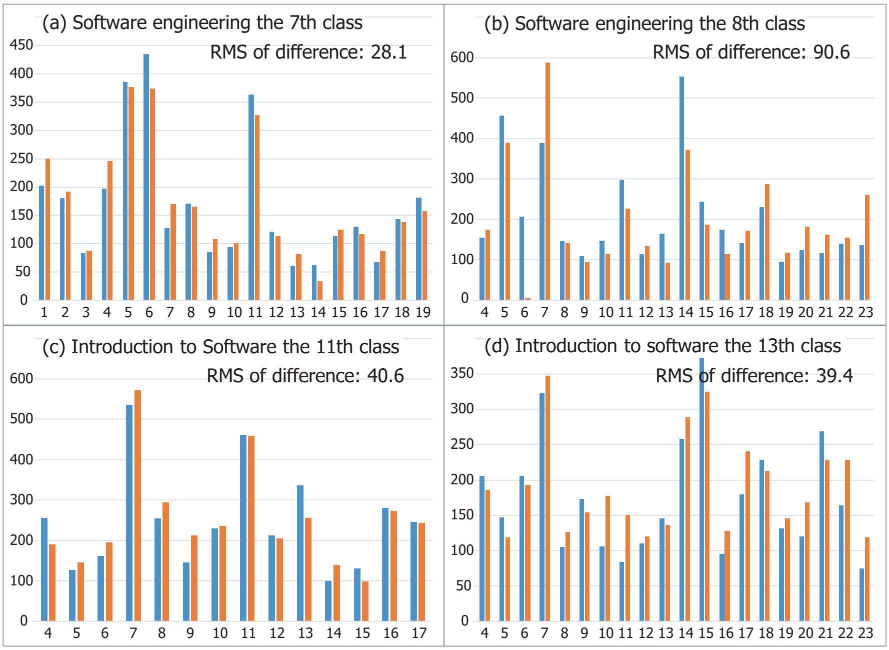


Fig. 14.10 Teacher's explanation time

Figure 14.10 shows the teacher's explanation time on each page of each class. The horizontal axis is the page number, and the vertical axis is the explanation time within the page.

In the comparison under the same conditions (a, c), there was no significant difference in the teacher's explanation time. The root mean square of the difference is 28, 1, and 40.6, respectively. On the other hand, when the fill-in-the-blank rate was hidden (b), the difference in explanation time grew larger. The root mean square of the difference is 90.6. In particular, there was a large difference between the exercise pages on pages 7 and 14. In addition, when the time-series heat map was hidden (4), there was no significant difference from when it was displayed. The root mean square of the difference is 39.4.

From these results, it was confirmed that the teacher could ascertain the progress of the students by referring to the fill-in-the-blank rate and proceeded with the lessons accordingly. We also confirmed that the time-series heat map does not affect the teacher's explanation time. Verification with a large amount of data is a future task.

14.7.2 Interview Survey with Teachers

We interviewed two teachers in charge of the lessons about the functions of each mode. The results are described below.

1. Authoring mode

   It was easy to set a blank. However, it was troublesome to move the mouse to the blank setting button.

   I want a function to change the size after setting a blank.

   I had to recreate the animated page.
2. Presentation mode

   It was useful because I could decide the timing for moving to the next page by referring to the fill-in-the-blank rate.

   In a lesson without a fill-in-the-blank rate, progress through the material unknowingly speeds up, and at the end of the lesson, there is increased student demand to see the previous page again.

   In the lesson where the fill-in-the-blank rate is not displayed, the time-series heat map of the lesson attitude was referred to, and I accelerated the pace when the number of students who proceeded increased.

   It was good to be able to know the situation of the students by looking at the time-series heat map of the class attitude.

   The time-series heat map of the lesson attitude sometimes deviates greatly from what is perceived using the senses, so it is unreliable.

   Therefore, no particular action was taken based on the time-series heat map.
3. Analytics mode

   It is good to know where the students' attitude deteriorates in the class.

   I want a function such as displaying the worst 3 lessons that should be improved out of all 15 lessons.

   The reliability of the time-series heat map remains questionable.

As mentioned above, it was confirmed that this system is useful for improving lessons. Future tasks are to verify the validity of the classification of class attitudes, improve the time-series heat map, and improve the user interface of each mode.

14.7.3 Questionnaire Survey of Students

We conducted a questionnaire survey of students to confirm the usability and usefulness of the system. The results are shown in Fig. 14.11.

From this result, it was confirmed that the speed of the lesson was adjusted more appropriately when the system was used, and the students felt that it was easier to concentrate on the lesson and to understand the contents of the lesson. Not many students answered that the class attitude of the analytics mode matched their feelings, and this needs to be improved.

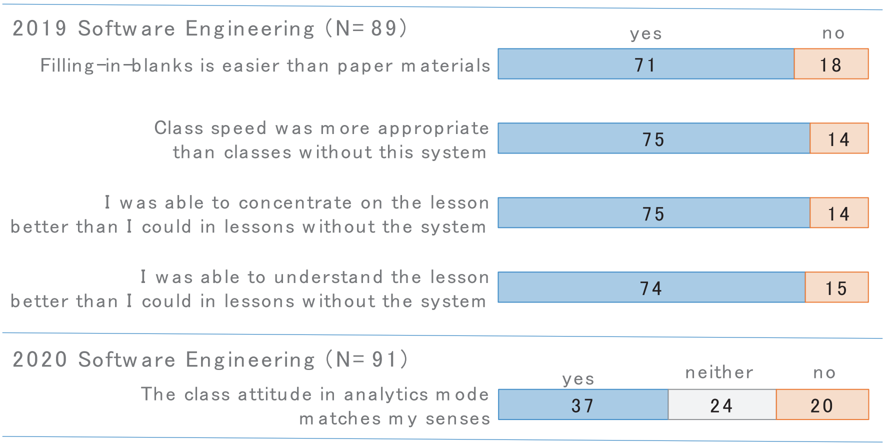


Fig. 14.11 Results of questionnaire survey of students

14.8 Conclusion

In order to help teachers understand student learning attitudes and improve their classes, the authors have developed a fill-in workbook system. This system provides an authoring mode in which teachers create teaching materials, a presentation mode in which teachers explain teaching materials while referring to the students' learning attitude, a browsing mode in which students refer to teaching materials while filling in the blanks, and an analytics mode to support a retrospective review of teachers and students.

By using the system, students are compelled to interact with the system, and the operation logs of inactive students who do not actively participate in learning activities by, for example, underlining and writing notes can be acquired. This allows teachers to adjust the pace of their lessons when more students are ahead or behind. In addition, when the number of students who are not participating increases, it is possible to take actions to encourage the concentration of students, such as having them discuss with each other and imposing exercises. Furthermore, after class, the teacher can check the transition of the learning attitude of the entire class and identify the points for improvement. Students can look back on their class attitude.

We applied this system to an actual class attended by about 100 people and confirmed that (1) it was possible to easily create fill-in-the-blank electronic teaching materials in the authoring mode, (2) the teacher can ascertain the learning attitude of the whole class in real time and adjust the lesson speed in the presentation mode, (3) the student can easily fill in the blanks in the browsing mode, (4) the analytics mode can be used for the teacher's and the students' retrospective review after the lecture. Future tasks are to verify the validity of the classification of class attitudes and improve the time-series heat map and analytics mode.

Acknowledgments We would like to thank Associate Professor Yashiro Sudo and the students of the Department of Information and Computer Science, Faculty of Information Technology, Kanagawa Institute of Technology, for their cooperation in the classes using this system and the subsequent interviews and questionnaires.

This work was supported by JSPS KAKENHI Grant Number JP20K03078.

References

Abe, K., et al. (2018). Lecture support system using digital textbook for filling in blanks to visualize student learning behavior. *International Journal of Education and Learning Systems, 3*, 138–144.

Abe, K., et al. (2019). Analysis of learning behavior logs based on a fill-in workbook system - Toward real-time learning analytics. *International Journal of Education and Information Technologies, 13*, 154–159.

Abe, K., et al. (2020a). Learning attitude analysis based on operation log of fill-in workbook system. In *Proceedings of IIAI 9th International Congress on Advanced Applied Informatics,* Online, to appear.

Abe, K., et al. (2020b). Prototyping a retrospective review support function for a fill-in workbook system. In *Proceedings of the 17th International Conference on Cognition and Exploratory Learning in Digital Age (CELDA 2020)* (pp. 243–249).

Diana, M., et al. (2019). A visual dashboard to track learning analytics for educational cloud computing. *Sensors, 19*, 2952. https://doi.org/10.3390/s19132952

Ferguson, R. (2012). Learning analytics: Drivers, developments and challenges. *International Journal of Technology Enhanced Learning, 4*(5/6), 304–317.

Fu, X., et al. (2017). Real-time learning analytics for C programming language courses. In *Proceedings of the Seventh International Learning Analytics & Knowledge (LAK '17)*, Vancouver (pp. 280–288).

Kumar, K., & Vivekanandan, V. (2018). Advancing learning through smart learning analytics: a review of case studies. *Asian Association of Open Universities Journal, 13*(1), 1–12.

Reynolds, G. (2005). *Presentation Zen: The "Monta Method"*. http://presentationzen.blogs.com/presentationzen/2005/10/the_monta_metho.html.

Schwartz, C., (1991). An academic adventure. http://socrates.berkeley.edu/~schwrtz/Minute.html.

Shimada, A., et al. (2018). Real-time learning analytics system for improvement of on-site lectures. *Interactive Technology and Smart Education, 15*(4), 314–331.

Shimada, A., & Konomi, S. (2017). A lecture supporting system based on real-time learning analytics. In *Proceedings of CELDA2017*, Vilamoura, Algarve (pp. 197–204).

Verbert, K., et al. (2020). Learning analytics dashboards: The past, the present and the future. In: *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge (LAK '20)*, Cyberspace (pp. 35–40).

Viberg, O., et al. (2018). The current landscape of learning analytics in higher education. *Computers in Human Behavior, 89*, 98–110.

Chapter 15
MAI Helper: Learning Support System for Time Management Skill Acquisition Using Learning Analytics

Hiroyuki Watanabe, Li Chen, Xuewang Geng, Yoshiko Goda, Atsushi Shimada, and Masanori Yamada

15.1 Introduction

Self-regulated learning (SRL) is an important factor in students' learning processes and the environment. Zimmerman et al. (1996) refer to self-regulation as the active involvement of learners in their learning process in metacognition, motivation, and behaviors.

Accordingly, these are effective perspectives for achieving learning performance and academic goals. With recent advances in information and communications technology, learning analytics can capture the data of learners' actual learning behaviors. Yamada et al. (2015, 2016a, 2017a, b, 2018) indicated that SRL awareness is related to learning behaviors. Chen et al. (2019a) also indicated that cognitive learning behaviors in reading learning materials enhance SRL awareness and learning performance. Yamada et al. (2015, 2016a, b) indicated that awareness of SRL promotes effectiveness, which leads to remembering the submission deadline for an assignment. Thus, learning analytics is a promising approach to investigate the factors of SRL at the behavioral level.

H. Watanabe (✉) · L. Chen · M. Yamada
Faculty of Arts and Science, Kyushu University, Fukuoka, Japan
e-mail: watanabe@artsci.kyushu-u.ac.jp

X. Geng
Graduate School of Human-Environment Studies, Kyushu University, Fukuoka, Japan

Y. Goda
Research Center for Instructional Systems, Kumamoto University, Kumamoto, Japan

A. Shimada
Faculty of Information Science and Electrical Engineering, Kyushu University, Fukuoka, Japan

D. Ifenthaler et al. (eds.), *Orchestration of Learning Environments in the Digital World*, Cognition and Exploratory Learning in the Digital Age,
https://doi.org/10.1007/978-3-030-90944-4_15

In particular, learning data extracted and collected by the learning analytics approach present students' learning behaviors, which can be related to their learning skills. Therefore, it is necessary to identify the basic learning skills that can be measured by learning log data based on previous research. Learning skills can be divided into three types, based on a previous referenced review (Watanabe et al., 2020a, b).

The first is "information processing," which includes the competencies associated with acquiring, recording, organizing, synthesizing, remembering, and using information (Hoover & Patton, 1995).

The second is "skills of learning attitudes and habits of learning," which refers to the abilities, habits, understanding, and attitudes necessary to achieve the goal of learning (Cottrell, 2019). Tanigawa et al. (2014) described learning skills as "the comprehensive skills needed to learn effectively in higher education institutions, some related to learning skills and others to learning attitudes and habits." Specifically, this refers to time management, taking classes (how to learn to prepare, review, and prepare for exams, how to take lecture notes), reading (how to find literature and the skill of effectively reading it), and writing (summarizing ideas, writing clearly).

The third type is "study strategy," which includes time management and time planning, test anticipation and preparation, and reading and writing strategies as the characteristics of successful students (Zimmerman et al., 1996). Gettinger and Seibert (2002) used an information-processing framework to divide learning skills strategies into four categories: (1) repetition- or rehearsal-based skills, (2) procedural study skills, (3) cognitive-based study skills, and (4) metacognitive skills.

Finally, regarding the assessment of basic learning skills, learning skills inventories were mainly used to assess the achievement of learning skills. Typical examples of learning skills inventories include the learning and study strategies inventory (LASSI), the approaches and study skills inventory for students (ASSIST), and the Dennis Congos study scale inventory (DCSSI). For example, the DCSSI, created by Dennis H. Congos from the University of Central Florida, is used primarily by universities in the United States. DCSSI comprises 6 scales, reading, note-taking, memory, test preparation, concentration, and time management; moreover, there are 51 subscales (Congos, 2011).

Although learning skills are classified into different categories, time management is a common element in all categories. Effective time management is an essential skill for students to implement their learning plans; it is also essential to achieve excellent learning outcomes (Zimmerman et al., 1996).

Thus, time management is regarded as an essential part of learning skills. For example, even if students have prepared and adequately made plans for their work, they will still have difficulties completing the work as planned without proper time management, since it inhibits their self-efficacy. Hence, this study designs a learning analytics-based system that helps acquire learning time management skills. Time management is an essential learning skill for enhancing learning performance.

Students gradually develop study habits and subsequently feel that having a learning plan is beneficial for their learning (Goda et al., 2009). Time-related learning patterns affect learning performance (Goda et al., 2015).

Britton and Tesser (1991) examined the hypothesis that college students' performance is predicted by time management practices. They found that time management practices could positively affect college performance (academic performance).

How is time management skills enhanced? Talib and Sansgiry (2012) demonstrated that factors such as time management and test anxiety were significantly related to students' academic performance. Furthermore, Misra and McKean (2000) reported that time management behaviors have a greater buffering effect on academic stress than leisure satisfaction activities. The enhancement of time management is related to awareness of time.

Yamada et al. (2015) indicated that learners who have high timely awareness, reflected by both negative and positive self-concepts about learning, tended to have less procrastination awareness and had high learning performance; however, these studies depend on self-report data.

From the insights of the above studies, it is indicated that time management behaviors and practices show positive relationships with students' learning outcomes and positive effects on academic stress. Therefore, it is important to consider time management when teaching and learning.

15.2 The Relationships Between Time Management and Learning Analytics

The following section presents previous studies that investigated learning time management skills using a learning analytics approach. Many studies on time management skills were mainly conducted in educational psychology research, particularly as a part of SRL. However, learning analytics contributes to improving learning and educational environments using various data in psychological data and learning logs. For example, Okubo et al. (2016) investigated effective learning behaviors to predict learning performance using neural networks. Yin et al. (2019) indicated that reading previous research repeatedly on learning materials promotes learning performance. Shimada et al. (2015) indicated that learning outside of class enhances learning performance, from the perspective of learning analytics.

Tabuenca et al. (2015) used mobile devices as tools to facilitate SRL in online courses and measured the actual time spent in learning over 4 months. The results show that tracing time can improve time management skills.

Uzir et al. (2020) conducted an exploratory study using data collected from a graduate computer course to detect time management strategies based on actual students' learning activities. The results showed that time management strategies were significantly associated with academic performance. Thus, to measure time management in terms of student behavior, a learning analytics approach could be very useful. Nevertheless, these studies did not specify the process of acquisition of time management skills.

Schumacher and Ifenthaler (2018) investigated students' expectations of the features of learning analytics systems and their willingness to use these features for learning. The five most commonly accepted learning analytics features by students were (1) the deadline reminder function, (2) the function of revising learning based on the previous semester outcomes, (3) the function of receiving prompts with self-assessment questions with just-in-time feedback, (4) the function of receiving feedback on assignments created in an online learning environment, and (5) a learning recommendation function for the successful completion of the course.

This study considers the students' perspectives (e.g., students' expectations). Using the learning analytics functions listed above, students can monitor their learning during the entire learning process. In particular, it was found that the deadline reminder function is an important function for students to organize their SRL and time management activities. These previous studies have shown that the learning analytics approach is effective in measuring time management skills. The studies also showed that students wanted to manage their learning activities. Based on the results of the above studies, we designed and developed a system "MAI (Management and Active involvement of Learning) Helper" that enables users to manage time to monitor their learning activities.

15.3 Design, Development, and Function of "The MAI Helper"

15.3.1 Design to Visualize Time Management

First, the system focuses on visualization and description of information such as time spent, access to resources, progress, and comparison with other students in the same class. Therefore, this application will allow students to understand their learning status, especially in areas where they cannot manage their schedules well. Furthermore, promoting metacognition in time management will improve knowledge retention.

15.3.2 System Development and Functions

Regarding the functions of the MAI Helper, we developed the visualization of each graph, access time, schedule progress, reflection, comparisons between the students and their classmates, and integration with Bookroll (BR) and Metaboard (MB). BR refers to electronic textbooks, which can be used to record the use of digital lecture materials such as slides and notes (Ogata et al., 2015, 2017). MB is a learning analytics dashboard that aims to visualize processes and behaviors based on the learning log data of the BR system (Chen et al., 2019b, 2020; Lu et al., 2020). Data from the BR and MB systems can be obtained and visualized using the MAI Helper, as displayed in Fig. 15.1 (Watanabe et al., 2021).

Below the MAI Helper logo is the list of courses registered in the schedule since the access date. The students' names, periods, and subjects are displayed on the right. On the left side of the screen, users can select the day and the week. Accordingly, the deeper the color, the more access to the site. Users can register the contents of each schedule if they select the schedule tab. Regarding the "Bookroll," "Metaboard," and "Reflection" tabs on the left side, the darker the color, the more accessible they are. Students can also see their access date and time, the class access date and time, and the number of people in the class by hovering the cursor over the tab. On the right side, the schedule shows the access day, and the students can check the detailed contents by moving the cursor to edit or delete items. In addition, there is a reminder function, and if students check the box when entering the schedule, they will be notified by email at the desired date and time (Fig. 15.2). If users select the "Reflection" tab, they can reflect on their learning and make plans for the next time (Fig. 15.3).

In addition, the teacher-only functions include the registration of class information (Fig. 15.4) and downloading access logs (Zarzour et al., 2020) (Table 15.1). The background is colored in the morning, noon, evening, and night so that the user can intuitively understand the access time (Fig. 15.5). The specific access method was as follows:

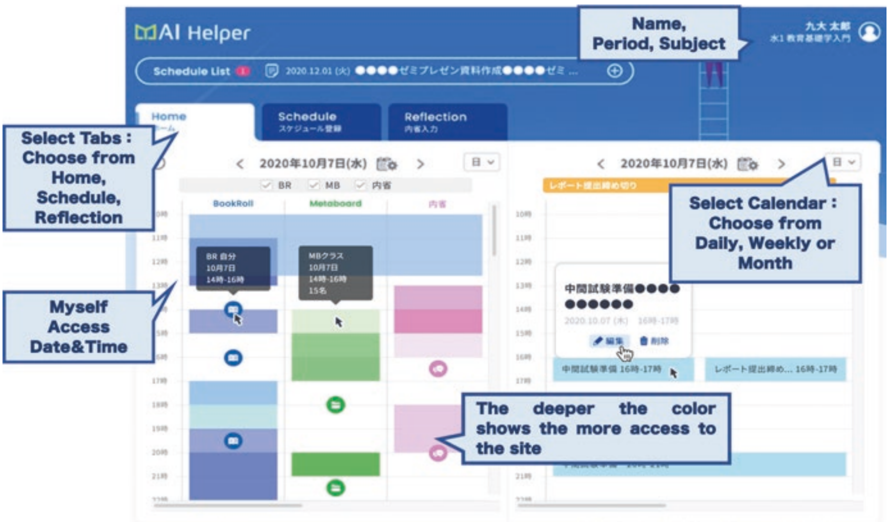


Fig. 15.1 The top screen of MAI Helper

Fig. 15.2 Schedule registration screen

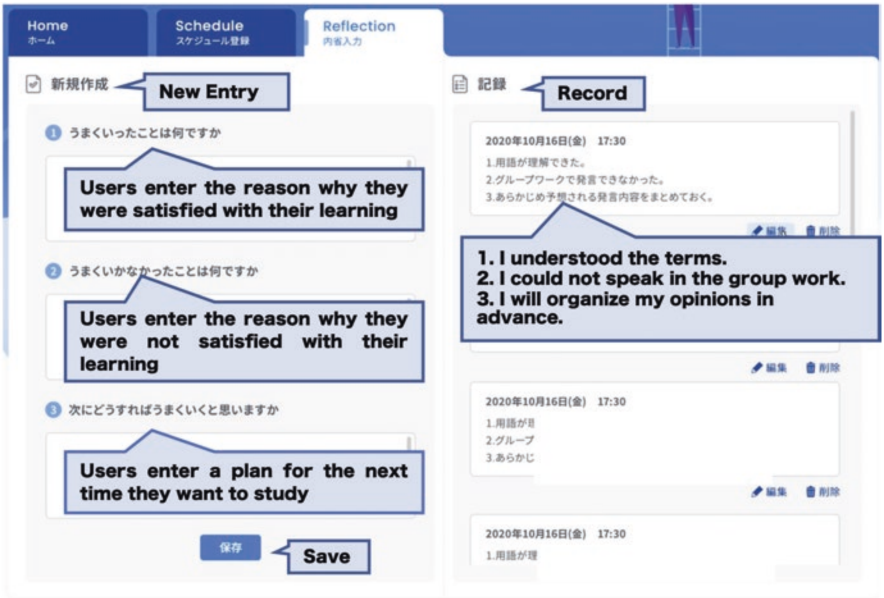


Fig. 15.3 Reflection screen

First, the students access the campus using a single sign-on system on a PC, tablet, etc. They then access the learning management system (LMS), and after the learning tools interoperability (LTI) linkage, they access the learning analytics

Fig. 15.4 Registration of class information

dashboard (LAD). The MAI Helper is then positioned as a type of learning dashboard (Fig. 15.6). The database used by the MAI Helper database is MySQL.

15.4 Formative Evaluation

Next, we conducted a formative evaluation before using the MAI Helper (Dick et al., 2001; Suzuki, 2002). A formative evaluation was originally designed to improve teaching materials.

MAI Helper, the subject of this evaluation, is not a teaching material; instead, it is a system that supports time management. However, by using the system beforehand, we can check for problems and use it as a material to determine whether it can withstand applications. Formative evaluation usually involves one-on-one evaluations, small-group evaluations, and field trials (Dick et al., 2001).

The MAI Helper is accessed online, and in this case, we conducted a one-on-one evaluation, the first stage of formative evaluation. Specifically, we used Zoom to check the usage status of the learners and measure their usage time. In other words, the evaluation was conducted in an environment that was almost the same as face to face. It also has the advantage of obtaining log data on usage, allowing us to correct simple errors in the system and explore the problems.

15.4.1 The Evaluation Procedure

The evaluation procedure was done as follows:

1. A simple manual was provided in advance.
2. On the day of the evaluation, after explaining the purpose and usage, the participants were asked to carry out the procedure.

Table 15.1 Example of schedule log

Student No.	Name	Update type	Schedule title	Start time	End time	Repeat setting	Repeat setting contents	E-mail notification setting: check box	E-mail notification setting: pull down	E-mail notification settings	E-mail notification time	Log making time
1375	Kyu Taro	Addition	test1	2021/1/10 10:00	2021/1/10 10:10	Yes	Everyday	Yes	Yes	Notification	2021/1/21 10:00	2021/1/10 10:10
1375	Kyu Taro	Addition	class	2021/1/12 13:00	2121/1/12 14:30	Yes	Everyday	No	No	No notification		2021/1/12 14:30
1375	Kyu Taro	Addition	report	2021/1/21 20:00	2021/1/25 21:30	Yes	Everyday	Yes	Yes	Notification	2021/1/25 10:00	2021/1/25 21:30
1375	Kyu Taro	Addition	library	2021/2/4 12:00	2021/2/4 16:30	Yes	Everyday	No	No	No notification		2021/2/1 14:57
1375	Kyu Taro	Addition	english	2021/2/5 15:00	2021/2/5 15:30	Yes	Everyday	No	No	No notification		2021/2/1 14:57

Fig. 15.5 Color coding by access time (from top to bottom: morning, noon, evening, night)

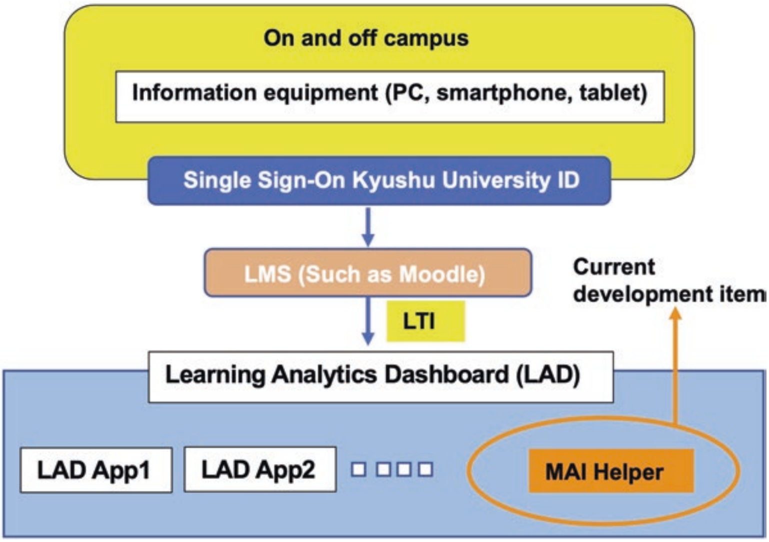


Fig. 15.6 Login to MAI Helper

A questionnaire was administered after the completion of the evaluation. MAI Helper can also be accessed using a tablet, but we chose a PC since all students have access to it.

First, we explain the formative evaluation. Next, we share our PC screens on Zoom and show the actual operation based on the following observation plan. After that, we ask the participants to share their operation screens on Zoom and ask them to perform the operation on their own while observing them. Table 15.2 displays the observation plan.

Table 15.2 Observation plan

Order #	Actions
1	Log in to Moodle for the Educational Information Technology class
2	Select Bookroll for this class and read ten pages
3	Return to the class Moodle
4	Select Metaboard
5	Select MAI Helper
6	Click on the "i" in the Home tab to see the meaning of the icons and heat map
7	Check the heat map display on BR, MB, MH in the Home tab
8	Click the icon next to the date to pop up the calendar
9	The display of the day, week, and month can be selected from the right side of the date
10	Select the Schedule tab
11	Select March 29 (Monday) from the calendar and press Enter(specific date for this time only)
12	Any time is fine, so click it, and the schedule registration screen will pop up
13	Enter "Report deadline" in "Enter title"
14	Check "All day"
15	Check "Notify me by email" and press the save button
16	"Report deadline has been created." pops up, so press the close button
17	The screen will switch to the schedule for March 29, 2021 (Monday), so check if the "Report deadline" is registered
18	You can select the display of the day, week, and month from the right side of the date
19	Select the Reflection tab
20	Write the reflection in the new creation field on the left and press save
21	Check if the description is reflected in the record on the right
22	Return to Home again and check the BR, MB, MH heat maps
23	Logout

15.4.2 The Evaluation Method

After the participants had completed these tasks, the next step was to conduct a questionnaire on the system's usability. Usability is considered a perspective in this evaluation. Nielsen (1994) showed five usability aspects: (1) learnability, (2) efficiency, (3) memorability, (4) errors, and (5) satisfaction. These characteristics consider not only ease of use but also the ease of accomplishing the task.

In the evaluation, we applied commonly used indicators to enable a quantitative evaluation. The main indicators of satisfaction are the system usability scale (SUS) (Brooke, 1996), computer system usability questionnaire (CSUQ) (Lewis, 1995), and questionnaire for user interface satisfaction (QUIS) (Chin et al., 1988).

Among these, we selected the SUS. This is because the template of the questions and the method of tabulation and evaluation are clearly defined. Even with small sample size, the SUS is more accurate than other questionnaires. It has a higher consistency among respondents (Tullis & Stetson, 2004). The SUS is characterized by ten items, with odd-numbered items being negative items and even-numbered items being positive items. The following response format was used (Table 15.3).

Table 15.3 SUS response format (Brooke, 1996; Sauro, 2011)

Strongly disagree 1	2	3	4	Strongly agree 5
○	○	○	○	○

Table 15.4 SUS items

1	I think that I would like to use this system frequently
2	I found the system unnecessarily complex
3	I thought the system was easy to use
4	I think that I would need the support of a technical person to be able to use this system
5	I found the various functions in this system were well integrated
6	I thought there was too much inconsistency in this system
7	I would imagine that most people would learn to use this system very quickly
8	I found the system very cumbersome to use
9	I felt very confident using the system
10	I needed to learn a lot of things before I could get going with this system

Afterward, the students were asked to fill out items on Moodle. The items in the SUS are displayed in Table 15.4.

We calculated the SUS scores as follows:

1. For odd numbers: Subtract 1 from the user's answer.
2. For even numbers: Subtract the user's answer from 5.
3. Total each user's answers and multiply the sum by 2.5.
4. This will convert your score from 0 to 100.

15.4.3 The Questionnaire Results

The results of the questionnaire, based on the SUS tabulation method, are displayed in Table 15.5.

Q1: SUS

The average score of the nine participants was 82.2 (SD = 12.4).

Next are the results of the free writing. There were three open-ended questions.

First, Q3 and Q4 are about the good points and improvements of the system. These questions categorize the responses according to the following three features of the system:

(a) Visualizing learning behavior
(b) Creating and checking schedules
(c) Creating reflections

Table 15.5 SUS score ($N = 9$)

Students	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Total	Conversion
A	4	4	4	4	4	4	4	4	3	4	39	97.5
B	4	3	4	3	3	3	4	4	4	0	32	80
C	1	4	4	3	3	3	4	4	4	4	34	85
D	4	4	4	0	4	4	3	4	4	1	32	80
E	4	4	4	3	3	3	3	4	3	4	35	87.5
F	4	3	4	1	4	4	3	4	4	3	34	85
G	2	4	4	3	4	4	4	4	4	3	36	90
H	3	2	3	0	2	2	1	4	3	0	20	50
I	3	4	4	4	2	4	3	3	4	3	34	85
Total	29	32	35	21	29	31	29	35	33	22	296	740
Average	3.2	3.6	3.9	2.3	3.2	3.4	3.2	3.9	3.7	2.4	32.9	82.2
SD	1.1	0.7	0.3	1.6	0.8	0.7	1.0	0.3	0.5	1.7	5.28	12.4

Q3: Please list three things that you find effective in this system.

(a) Regarding the visualization of learning behavior, seven out of nine students answered this question.

They answered that they could compare their own behavior as well as that of others. Regarding the benefits of comparison with others, some said that it gives them confidence in their own learning and motivation when they are not motivated.

Specific Items.

- "I can compare my learning behavior with other classmates." (student A)
- "I can see my access status and that of others." (student B).
- "It is easy to go back to the history of my use of Bookroll and Metaboard." (student I).

(b) All of the nine respondents responded to creating and checking schedules.

- From the responses, we can assume that the UI of the schedule function is excellent and that the search and reminder functions are well received.

- Specific Items.
- "The visualization of the class status and my own status make it easy to create my own schedule." (student B).
- "The UI for displaying, adding, and searching schedules is intuitive and easy to use." (student E).
- "Reminder emails can be sent before the scheduled date and time." (student H).
- "When you access the site, the HOME screen displays the day's information and schedule in an easy-to-understand manner." (student G).

(c) Six out of nine respondents answered about creating reflections.

This evaluation suggests that it is easy to acquire the habit of reflection.

Specific Items

- "The reflection tasks are basic, simple, and easy to use." (student B).
- "I can write reflections while looking at past content." (student C).
- "I can record my thoughts and reflections of the day." (student F).

Q4: Please list three things that you feel need to be improved in this system.

(a) Four out of nine students answered about visualization of learning behavior.

From the responses, students seem to not know the criteria for visualization. Since the MAI Helper is included in Metaboard, the same could be done. In addition, the number of times is displayed, and the color depth is in principle based on the number of accesses. These can be resolved by first explaining them in the manual. There were also suggestions for labeling BR, MB, and MH and for indicating the type of Metaboard used. We will try to improve these if possible.

Specific Items

- "Metaboard and MAI Helper records look the same, so it is hard to tell the difference." (student A).
- "I don't understand the meaning of the number of times BR, MB, and MH are displayed." (student D).
- "I don't understand the criteria used to determine the dark and light colors of BR, MB, and MH." (student D).
- "I don't understand the criteria used to determine the dark and light colors of BR, MB, and MH. I want to see the labels of BR, MB, and MH even when scrolling." (student G).
- "When BR, MB, and MH checkboxes are shown or hidden, only the ones shown should have their columns expanded for easier viewing." (student G).
- "If you can see which Metaboard is being used by others in the class, it will increase interest in the tool." (student C).

(b) Five of the nine students responded to the question about creating and checking schedules.

Here, suggestions were made regarding the behavior of the screen and buttons. These will be considered for improvement.

Specific Items

- "When entering data from the month or week, the screen returns to the day screen, making it difficult to check." (student I).
- "It would be better if there was a confirmation message before deleting." (student E).
- "When viewing a list of schedules, the ⊕ button has an additional meaning, so there is a gap between the intuitive understanding of the button and the actual behavior." (student F).
- "Letting users enter weekly or monthly goals is useful for planning and reflection." (student B).

(c) Creating reflection was answered by four out of nine respondents.

Here, there were suggestions about adding a reflection function and linking it with the schedule function. Since it needs to be linked with other functions, improvements will be considered.

Specific Items

- "Visualize the degree of completion of the schedule by linking the reflection function with the schedule function." (student B).
- "I would like to see a memo function added to reflection." (student C).
- "Add a memo function to the reflection function. – Make it possible to describe the reflection function separately for classes, schedules, etc." (student A).

Next, Q5 was a question about tips for time management. The answers were categorized into the following three items:

(a) Manage time by task.
(b) Clarifying learning goals.
(c) Use of monitoring and reminders.

These were answered by all nine respondents.

Q5: What are the things and actions you think are necessary to improve your "time management skills"?

(a) Managing Time by Task.

This means that learning time varies depending on the task, so it is necessary to create a plan to deal with this.

Specific Items

- "In addition to the schedule, manage the time, resources, and order of tasks." (student A).
- "Allow plenty of time before the deadline." (student C).
- "Always be aware of time." (student H).

(b) Clarify Learning Goals.

This means to be aware of learning goals and tasks and create a schedule according to those goals.

Specific Items

- "Plan the goals and specific actions and schedules for each stage according to the learning goals." (student B).

(c) Use of Monitoring and Reminders.

It means that to carry out one's plan, one must constantly monitor the level of achievement.

Specific Items

- "Frequent scheduling and reminders." (student G).
- "Constant monitoring of the plan, reflection, and adjustment of the plan."(student B)

From these free descriptions, we have learned how students manage their own study time, which will be used as a reference for future improvement.

15.5 Discussion

Based on these results, we would like to consider the relationship with the frequency of the questionnaire. According to Sauro (2011), the average SUS score for 500 different evaluations with over 5000 participants was 68. In other words, if the score is above 68, it is above average. However, it should be noted that the SUS score was not a percentage. The conversion of this score into percentiles and letter grades is presented in Table 15.6.

When the results were applied, the average score of all the participants was 82.2 (SD = 12.4), which is higher than the usual average score of 68. As shown in Table 15.6, the grade is A, and the percentile range is 90–95, which is considered to be a passable usability evaluation.

However, looking at the individual results, the items with a lower total score than the others are "Q.4: I think I need support for a technician to use this system" and "Q.10: I need to learn a lot about this system before I can start using it." Q.4 and Q.10 are said to be the ease of learning scales. Why are these scores so low?

Student B had a score of 0 for Q.4 and Q.10. The reason for this is that "I feel that this system has too many elements." Therefore, student B stated that "A detailed manual needs to be prepared." Furthermore, student H says, "Regarding the heat maps for BR, MB, etc., at Home tab, the meaning of the frequency is not clear." Similarly, student D with a score of 1 said, "I do not know how to use the heat maps for BR, MB, and MH."

Table 15.6 Curved grading scale for the SUS (Lewis & Sauro, 2017)

Grade	SUS	Percentile range
A+	84.1–100	96–100
A	80.8–84.0	90–95
A−	78.9–80.7	85–89
B+	77.2–78.8	80–84
B	74.1–77.1	70–79
B−	72.6–74.0	65–69
C+	71.1–72.5	60–64
C	65.0–71.0	41–59
C−	62.7–64.9	35–40
D	51.7–62.6	15–34
F	0–51.6	0–14

This could mean that they need guidance to use the system because there are many elements. A simple manual was provided to the participants before starting the evaluation, but it did not describe the meaning of the heat map or how to use it. This may need to be included in the manual. In addition, one way to help students understand the operation is to add videos.

In addition, looking at the overall score, student *H*'s score was 50, which is low compared to the other students (with an average of 82.2). Specifically, student H made detailed suggestions to make the system easier to use, such as: "If I want to add a schedule at the same time, I want to be able to enter and register it from the button of the registered time so that I can use it" and "In the schedule list on the top page, the round plus button is easily misunderstood as meaning 'add'." After the evaluation, student H stated that he had evaluated the system from the perspective of a product manager. In other words, the fact that the evaluation was done from a different perspective from that of the students is thought to be the reason for the low score. It should also be noted that these results are based on a limited amount of time.

15.6 Conclusions and Future Works

In this study, we examined previous research on learning skills and designed a new application to help students manage their learning time. The application is offered to both students and teachers. In the design of this application, the functions of graphs, calendars, schedules, and reflections are necessary for time management to aid the monitoring and management of learning plans. We also considered usability by minimizing the screen transitions.

We conducted a formative evaluation of the system with graduate students and found that the score was good, but found some issues.

First, for the visualization of learning behavior, the meaning of the number of times BR, MB, and MH are displayed and their colors should be informed.

Second, we improve the display and operation of the schedule buttons, enhance the reflection function.

At last, the following is a list of items necessary to improve time management skills based on the graduate students' descriptions:

(a) Managing time by task.
(b) Clarifying learning goals.
(c) Use of monitoring and reminders.

These can be achieved by using the functions of the MAI Helper.

However, these graduate students were majoring in educational technology and information technology, and their understanding and interest in the content was probably higher than that of the undergraduate student. There may have been some bias regarding the scores and comments. In the next phase, the MAI Helper should be evaluated in class.

After correcting the issues, the future works of the project will be to examine how to utilize the designed system and student access logs and how they relate to class design through practical experiments (Chen et al. 2019c).

Acknowledgments This research was supported by JST AIP Grant No. JPMJCR19U1, JSPS Kakenhi JP19H01716, JP20H01727, and JP21K18134, Japan.

References

Britton, B. K., & Tesser, A. (1991). Effects of time-management practices on college gades. *Journal of Educational Psychology, 83*(3), 405–410.

Brooke, J. (1996). SUS: a "quick and dirty" usability scale. In P. W. Jordan, B. Thomas, B. A. Weerdmeester, & A. L. McClelland (Eds.), *Usability evaluation in industry*. Taylor and Francis.

Chen, L., Goda, Y., & Yamada, M. (2019a). Factors investigation of learning behaviors affecting learning performance and self-regulated learning. In *Proceedings of IEEE Teaching, Assessment and Learning for Engineering 2019*, Yogyakarta, Indonesia (pp. 676–681). https://doi.org/10.1109/TALE48000.2019.9225926.

Chen, L. Lu, M., Goda, Y. & Yamada, M. (2019b). Design of learning analytics dashboard supporting metacognition. In *Proceedings of 16th International Conference Cognition and Exploratory Learning in Digital Age (CELDA 2019)* (pp. 175–182).

Chen, L., Yoshimatsu, N., Goda, Y., Okubo, F., Taniguchi, Y., Oi, M., Konomi, S., Shimada, O., & H. & Yamada, M. (2019c). Direction of collaborative problem solving-based STEM learning by learning analytics approach. *Research and Practice in Technology Enhanced Learning, 14*(1), 1–28.

Chen, L., Lu, M., Goda, Y., Shimada, A., & Yamada, M.(2020) Factors of the use of learning analytics dashboard that affect metacognition. In *Proceedings of 17th International Conference Cognition and Exploratory Learning in Digital Age (CELDA 2020)* (pp. 295–302).

Chin, J. P., Diehl, V. A., & Norman, K. L. (1988). Development of an instrument measuring user satisfaction of the human-computer interface. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 213–218).

Congos, D. H. (2011). *Starting out in community college*. McGraw-Hill.

Cottrell, S. (2019). *The study skills handbook* (5th ed.). Red Globe Press.

Dick, W., Carey, L., & Carey, J. O. (2001). *The systematic design of instruction* (6th ed.). Longman.

Gettinger, M., & Seibert, K. (2002). Contributions of study skills to academic competence. *School Psychology Review, 31*(3), 350–365.

Goda, Y., Matsuda, T., Yamada, M., Saito, Y., Kato, H., & Miyagawa, H. (2009). Ingenious attempts to develop self-regulated learning strategies with e-learning: Focusing on time-management skill and learning habit. In T. Bastiaens, et al. (Eds.), *Proceedings of World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education 2009*, Chesapeake, (pp. 1265–1274).

Goda, Y., Yamada, M., Kato, H., Matsuda, T., Saito, Y., & Miyagawa, H. (2015). Procrastination and other learning behavioral types in e-learning and their relationship with learning outcomes. *Learning and Individual Differences, 37*, 72–80. https://doi.org/10.1016/j.lindif.2014.11.001

Hoover, J. J., & Patton, J. R. (1995). *Teaching students with learning problems to use study skills: A teacher's guide*. Pro-Ed.

Lewis, J. R. (1995). IBM computer usability satisfaction questionnaires: psychometric evaluation and instructions for use. *International Journal of Human-Computer Interaction, 7*(1), 57–78.

Lewis, J. R., & Sauro, J. (2017). Can I leave this one out? The effect of dropping an item from the SUS. *Journal of Usability Studies,*13(1), 38–46.

Lu, M., Chen, L., Goda, Y., Shimada, A., & Yamada, M. (2020). Development of a learning dashboard prototype supporting meta-cognition for students. In *Companion Proceedings 10th International Conference on Learning Analytics & Knowledge (LAK20)*, Frankfurt (online), Germany (pp. 104–106).

Misra, R., & McKean, M. (2000). College students' academic stress and its relation to their anxiety, time management, and leisure satisfaction. *American Journal of Health Studies, 16*(1), 41–51.

Nielsen, J. (1994). *Usability engineering*. Morgan Kaufmann.

Ogata, H., Yin, C., Oi, M., Okubo, F., Shimada, A., Kojima, K., & Yamada, M. (2015). E-book-based learning analytics in university education. In *Proceedings of the 23rd International Conference on Computers in Education (ICCE 2015)* (pp. 401–406).

Ogata, H., Taniguchi, Y., Suehiro, D., Shimada, A., Oi, M., Okubo, F., Yamada, M. & Kojima, K. (2017). M2B system: A digital learning platform for traditional classrooms in university. In *7th International Learning Analytics & Knowledge Conference (LAK17): Practitioner Track*, Vancouver, Canada (pp. 155–162).

Okubo, F., Hirokawa, S., Oi, M., Yin, C., Shimada, A., Kojima, K., Yamada, M., & Ogata, H. (2016). Learning activity features of high performance students. In *Proceedings of the 1st International Workshop on Learning Analytics Across Physical and Digital Spaces (Cross-LAK 2016)* (pp. 28–33).

Sauro, J. (2011). *A practical guide to the System Usability Scale*. Measuring Usability LLC.

Schumacher, C., & Ifenthaler, D. (2018). Features students really expect from learning analytics. *Computers in Human Behavior, 78*, 397–407.

Shimada, A., Okubo, F., Yin, C., Kojima, K., Yamada, M., & Ogata, H. (2015). Informal learning behavior analysis using action logs and slide features in e-textbooks. In *Proceedings of IEEE ICALT 2015* (pp. 116–117).

Suzuki, K. (2002). *Instructional design manual*. Kitaoji-shobo.

Tabuenca, B., Kalz, M., Drachsler, H., & Specht, M. (2015). Time will tell: The role of mobile learning analytics in self-regulated learning. *Computers & Education, 89*, 53–74.

Talib, N., & Sansgiry, S. S. (2012). Determinants of academic performance of university students. *Pakistan Journal of Psychological Research*, 265–278.

Tanigawa, H., Ishige, Y., Tsukayama, J., Yamazato, K., Nagao, K., Shimosaka, T., & Kabeya, K. (2014). *Peer tutor training*. Nakanishiya.

Tullis, T. S., & Stetson, J. N. (2004). A comparison of questionnaires for assessing website usability. In Usability professional association conference (Vol. 1, pp. 1–12).

Uzir, N. A., Gašević, D., Matcha, W., Jovanović, J., & Pardo, A. (2020). Analytics of time management strategies in a flipped classroom. *Journal of Computer Assisted Learning, 36*(1), 70–88.

Watanabe, H., Chen, L., Geng, X., Goda, Y., & Shimada, A. (2020a). Design of the time management skills acquisition system using learning analytics. In *Proceedings of 17th International Conference Cognition and Exploratory Learning in Digital Age (CELDA 2020)* (pp. 67–73).

Watanabe, H., Chen, L., Shimada, A., & Yamada, M. (2020b). Toward identification of learning skills from learning analytics approach: A research review. In *2020 IEEE International Conference on Teaching, Assessment, and Learning for Engineering (TALE)* (pp. 807–812). IEEE.

Watanabe, H., Chen, L., Goda, Y., Shimada, A., & Yamada, M. (2021). Development of a time management skill support system based on learning analytics. In *Companion Proceedings of the 11th International Conference on Learning Analytics & Knowledge LAK20* (pp. 241–249).

Yamada, M., Goda, Y., Matsuda, T., Kato, H., & Miyagawa, H. (2015). The relationship among self-regulated learning, procrastination, and learning behaviors in blended learning environment. In *12th International Conference on Cognition and Exploratory Learning in Digital Age 2015 (CELDA 2015)* (pp. 67–74).

Yamada, M., Okubo, F., Oi, M., Shimada, A., Kojima, K., & Ogata, H. (2016a). Learning analytics in ubiquitous learning environments: Self-regulated learning perspective. In *The Proceedings of the 24th International Conference on Computers in Education (ICCE 2016)* (pp. 306–314).

Yamada, M., Goda, Y., Matsuda, T., Saito, Y., Kato, H., & Miyagawa, H. (2016b). How does self-regulated learning relate to active procrastination and other learning behaviors? *Journal of Computing in Higher Education, 28*(3), 326–343.

Yamada, M., Oi, M., & Konomi, S. (2017a). Are learning logs related to procrastination? From the viewpoint of self-regulated learning. In *Proceedings of 14th International Conference on Cognition and Exploratory Learning in the Digital Age (CELDA 2017)* (pp. 3–10). IADIS Press.

Yamada, M., Shimada, A., Okubo, F., Oi, M., Kojima, K., & Ogata, H. (2017b). Learning analytics of the relationships among self-regulated learning, learning behaviors, and learning performance. *Research and Practice in Technology Enhanced Learning, 12*, 13.

Yamada, M., Shimada, A., Oi, M., Taniguchi, Y., & Konomi, S. (2018). BR-Map: Concept map system using e-book logs. In *International Conference on Cognition and Exploratory Learning in Digital Age 2018 (CELDA 2018)* (pp. 248–254).

Yin, C., et al. (2019). Exploring the relationships between reading behavior patterns and learning outcomes based on log data from e-books: A human factor approach. *International Journal of Human-Computer Interaction, 35*(4–5), 313–322.

Zarzour, H., Bendjaballah, S., & Harirche, H. (2020). Exploring the behavioral patterns of students learning with a Facebook-based e-book approach. *Computers & Education, 156*, 103957.

Zimmerman, B. J., et al. (1996). *Developing self-regulated learner: Beyond achievement to self-efficacy*. American Psychological Association.

Chapter 16 The Development and Validation of a Concept Mapping Emotions Scale (CM-ES) for University Students

Xinyuan Li, Yanyan Duan, Tong Lu, and Guoqing Zhao

16.1 Introduction

Concept map is a knowledge organization and representation tool that uses nodes to represent concepts and links to represent relationships between concepts (Novak, 1984). As a thinking visualization tool, concept map has similar functions to cognitive tools (Jonassen, 1992). It can help students focus on challenging tasks by reducing their low-level cognitive load (Jonassen, 1992; Lajoie & Derry, 1993). Many studies have shown that concept map can promote students' critical thinking and other higher-level thinking skills (Hwang et al., 2020). Researches also show that concept map has an impact on students' learning motivation and academic performance (Chiou et al., 2020). As an external tool for learners, the concept map may add the cognitive load to learners and affect students' emotions and willingness to use concept map (Machado & Carvalho, 2020).

Emotion is a complex experience of consciousness, bodily sensation, and behavior. Academic emotion refers to the emotional experience of students in the learning process, which is closely related to academic motivation and self-awareness (Henderson & Milstein, 1996). It can not only affect students' learning (Chiang & Liu, 2014) but also affect students' self-system opening and closing. When the self-system is closed, the cognitive system and the metacognitive system are also difficult to function (Marzano & Kendall, 2006). Studies have shown that students' learning is deeply affected by emotions, which are related to specific learning situations and learning behaviors (Govaerts & Grégoire, 2008). Therefore, increasing researchers have begun to concern the emotional experience of students in different learning situations.

X. Li · Y. Duan · T. Lu · G. Zhao (✉)
Beijing Normal University, Beijing, China
e-mail: guoqingzh@bnu.edu.cn

D. Ifenthaler et al. (eds.), *Orchestration of Learning Environments in the Digital World*, Cognition and Exploratory Learning in the Digital Age,
https://doi.org/10.1007/978-3-030-90944-4_16

However, research related to concept maps seldom focuses on students' emotions when using concept maps to learn. Therefore, it is very necessary to understand students' emotions for concept mapping. Although there are many scales that can measure students' academic emotions (Govaerts & Grégoire, 2008), the general academic emotions scales cannot accurately represent the emotions in the specific situation of applying concept maps (Chiang & Liu, 2014). This study intends to develop a concept mapping emotions scale by revising the Student Academic Emotion Scale and tailoring it to the context of the concept mapping and test the reliability and validity of the scale. Compiling a scale to measure the application emotion of concept maps can not only help researchers understand the emotions of students when concept mapping but also provide a theoretical basis for the application of concept maps in teaching.

16.2 Literature Review

16.2.1 The Role of Concept Mapping in Instruction

Concept maps are usually used as cognitive tools or evaluation tools in teaching. Research on concept maps as a cognitive tool mainly focuses on the impact of concept maps on student learning. For example, Hwang et al. (2020) found that the problem-producing method based on concept maps can improve students' academic performance, critical thinking tendency, and the quality of problem-producing. Chiou et al. (2020) investigated the impact of structured computer-assisted collaborative concept mapping (SCACCM) on students' academic performance when used as a classroom aid for flipped classroom teaching (FCT); the results showed that this method can improve students' academic performance and motivation. When used as an evaluation tool, it usually includes the evaluation of the concept map and the evaluation of the students' knowledge construction through the concept map. For example, Cañas et al. (2015) elaborated the criteria and scoring method of a good concept map. In the study of Gracia-Moreno et al. (2017), the concept map was used as a tool to evaluate student learning to explore the use of private digital workspace and public workspace by French high school students. There are also some studies on the application of concept maps in teaching methods. For example, Tseng (2020) compared the differences in the impact of constructing concept maps and the filling concept maps for high school students in Taiwan. The results told that the critical thinking performance of the students in the concept map construction group is better than the concept map filling group.

However, concept maps are not useful in all situations, nor are they useful for all students. Immelman's (2018) research results indicated that concept mapping does not promote the conceptual understanding of South African accounting students. This is in contradiction with the research results that concept maps can promote the understanding of physics concepts in tenth-grade students (Farrokhnia et al., 2019).

Conflicting research results suggest other factors that influence the application of concept maps, such as students' cognitive factors, emotions, and teaching methods. A study found that concept mapping has less effect on students with higher prior knowledge level, so the mastery of prior knowledge of students may be a factor that affects the use of concept maps. At the same time, the study also pointed out that the length of time that students use concept maps and their mastery of concept maps before experiments may have an impact on the promotion of concept maps to learning (Brandt et al., 2001).

The teaching method is also an important factor affecting concept mapping. The latest research has begun to focus on the strategy of using concept maps. Liang et al. (2021) developed an alternate reality game with a mobile concept mapping-based questioning approach, which can promote effective learning of students. Some researchers paid attention to the impact of augmented reality technology on the application of concept maps and found that combining augmented reality and multidimensional concept maps would improve students' learning satisfaction, learning effect, and cognitive load (Chou et al., 2021). Roessger et al. (2018) explored whether exercises, feedback, and relational frameworks help to improve students' concept mapping skills in the long term. The results of the research indicated that repeated practice and feedback improved the structural quality of students' concept maps and increased the number of autobiographical explanations in these concept maps. Furtado et al. (2019) described the Airmap interface. It uses automatic layout management and space separation to reduce the cognitive load in the concept mapping process. That improves the effectiveness of the concept map application.

Emotions and cognition influence each other in the learning process. Concept maps can be used as cognitive tools to assist students in learning. Emotion may be a factor that affects the application effect of concept mapping. Researchers pointed out that some students show a certain resistance to the concept map (Brandt et al., 2001). Concept maps can promote meaningful learning for learners, but it is a difficult task to inspire students to change from rote learning practice to meaningful learning mode.

Many researchers have studied the reasons influencing the application of concept maps from the perspective of cognition and teaching methods, but few studies have elaborated from the perspective of emotions. That may be due to the lack of tools for measuring the emotions of concept maps. The development of a scale to measure the application emotion of concept maps can not only help researchers understand students' emotions when utilizing concept maps but also provide a theoretical basis for the application of concept maps in teaching.

16.2.2 The Influence of Emotion on Student Learning

Emotions are people's psychological and behavioral responses to objective things (Cannon, 1987; Damasio, 2005; James, 1884). Whether the object can meet the needs of the subject and the degree of satisfaction directly determines emotion. The widely accepted emotion classification is based on a two-dimensional model of arousal and valence (Reisenzein, 1994). Arousal refers to the strength of the emotional state, while valence divides emotions into positive emotions, neutral emotions, and negative emotions. Many studies have shown that there is a close relationship between students' emotions and learning. For example, some scholars have found that positive emotions can improve the efficiency and effect of learning. In contrast, negative emotions can reduce the efficiency and effect of learning (Tauber et al., 2017). This article combs the existing research and finds that emotions are mainly born through the cognitive activities of learning and have an impact on them. By this way, the emotions affect students' learning efficiency and learning effects (John et al., 2010; Roseman, 1984; Scherer et al., 2001).

Neuroscience (Damasio, 2005) had shown that emotion affects cognition, because the neural basis of emotion and cognition are consistent. In general, the influence of emotion on learning is reflected in the fact that emotion is the initiator, organizer, and maintainer of psychological activities. It affects human perception, interpretation of events, motivation, attention, memory, and cognitive activities such as decision-making ability (Wang, 2007).

Attention is the most initial step in the cognitive processing. In terms of attention, emotion will affect not only the allocation of attention resources but also the selectivity and continuity of attention. Negative emotions have a great influence on the allocation of attention resources, and positive emotions also help to improve students' selective attention and sustained attention (Gasper & Clore, 2002; Rowe et al., 2007). For example, Schmitz find that the individual's attention span in positive emotional states will expand with functional magnetic resonance imaging (fMRI) technology (Schmitz et al., 2009).

Memory is the recognition, retention, reappearance, or recognition of things by the human brain, which is the basis for advanced mental activities such as thinking and imagination. In terms of memory, researchers found that emotions mainly affect working memory. For example, the processing efficiency theory proposed by Eysenck and Calvo (1992) has explained that negative emotions will occupy part of working memory resources and reduce the efficiency of cognitive operations. Fredrickson's expansion-construction theory explains that positive emotions have instantaneous expansion functions and long-term construction functions (Fredrickson & Branigan, 2005). Emotions can affect people's cognition and execution ability. In addition, emotions are diffuse and would further affect people's physical, intellectual, and social adaptation (Foster & Lloyd, 2007).

Execution control mainly includes three functions: behavior inhibition, task conversion, and information update, which are necessary guarantees for us to complete learning activities. Learning itself is also affected by emotions, even the product of emotions. Emotions affect people's cognition and behavioral decision-making and can promote and inhibit learning.

The influence of emotion on learning is finally manifested on learning efficiency and learning effect. At present, there have been a large number of studies on the influence of emotions on learning efficiency and effectiveness, and many explain the reason from the aspect of cranial nerves. For example, the *Dopamine Handbook* explains that positive emotion will gradually increase the dopamine transmitter in the peripheral cortex. Transmitters can strengthen people's thinking flexibility to a large extent and enable people to have stronger response capabilities (Iversen et al., 2009). Therefore, a good mood can improve people's learning efficiency. Many scholars have adopted empirical research methods and found that positive emotions can promote learning efficiency and improve learning effects. In addition, good emotions could enable individuals and groups to have more concentrated attention and greater active participation in the process of work and study (Estrada et al., 1994). Also, they could improve the trust of individuals or groups (Myers & Tingley, 2017). Totally, individuals and groups have good interpersonal relationships if reducing work or study pressure, thereby improving learning efficiency.

It should be noted that the conclusions drawn from different studies may differ due to different measurement methods, research methods, learning tasks, and other factors. Some research results show that positive emotions may not necessarily improve the effect of learning. For example, Oaksford et al. (1996) found that neither positive nor negative emotions can interfere with reasoning. Other studies have found that happiness can promote simple reasoning and interfere with moderately difficult reasoning, neutral emotions have a better effect on moderately difficult reasoning, while sadness can influence any reasoning (Forgas, 1995). In addition, the specific context of emotions will also change the impact of emotions on learning (Götz et al., 2003).

16.2.3 Measurement of Emotions

One of the research fields on emotion is the measurement of emotion. Emotion is a multidimensional structure and multicomponent compound process, including internal experience, external expression, and physiological activation. It's manifestations are divided into explicit emotions and implicit emotions (Strongman, 2003). Among them, the measurement of explicit emotion is more convenient, because it is the most intuitive and direct expression of emotion. So, this article will introduce the measurement method of explicit emotion in detail.

With the advancement of cognitive neuroscience research technology, the measurement methods of explicit emotions have been greatly expanded, such as physiological signal detection, voice signal detection, facial expression recognition, body posture measurement, and so on. These technical methods are more scientific than self-reported measurement methods. They can not only collect information on emotional state in time but also reflect the entire process of emotional changes. However, such a measurement method has a heavy dependence on technology, and it cannot be used in this article due to the restriction of experimental conditions.

An earlier and widely influential measure of explicit emotion is self-report measurement. The three most basic questions about emotion measurement are timeliness, the background of emotion measurement, and the reliability and validity of emotion measurement. First of all, emotion is a comprehensive process that changes over time. The timeliness of measurement is an important factor affecting the accuracy of measurement results. As Robinson and Clore (2002) said, emotional self-reports should be based on current emotional experience based on emotional reports. Secondly, the individual's emotions are affected by specific scenarios, different task activities and emotions influence each other, and there are individual differences. Finally, the self-report scale must have good reliability and validity to ensure the reliability of the results. Commonly used self-reported measurement scales include ISO Project Mind Mirror Adjective List (MACL), Multiple Affective Adjective Checklist (MAACL), etc. The MACL has not been officially published so its authority is difficult to guarantee. MAACL with 132 items is too long and cumbersome to implement, and it only measures the three negative emotions of loss, anxiety, and hostility. The dimensions measured by PANAS include not only negative emotions but also positive emotions. There is a total of 20 five-point scoring items. The comprehensiveness and indirectness of the measurement are more applicable than the previous two measurement methods. However, the PANAS scale does not reflect the emotional state of a specific situation, so emotional scales in specific fields such as the academic mood scale have emerged in the academic community. The Academic Emotion Scale (AES) compiled by Govaerts and Grégoire (2008) contains six kinds of emotions related to academic work, and each emotion is represented by multiple items. The scale comprehensively measures multiple dimensions of emotion; its compilation process is standardized and it has good reliability and validity. So the Academic Emotion Scale developed by Govaerts and Grégoire (2008) was translated in Chinese and adapted to form the CM-ES in this study.

16.2.4 The Present Study

Among the existing studies, the research on concept maps has focused on testing the validity of concept maps in different teaching situations and using concept maps to evaluate students' academic performance. Few studies have focused on students' emotions when using concept maps. Machado and Carvalho (2020) found that students' frustration with concept maps can be reduced by teachers' counseling or peer discussions. However, there is still a lack of scales that can comprehensively measure various emotional states of students when applying concept maps, which may influence students' study.

Measuring students' emotions toward concept map can help teachers understand students' emotional experience in teaching practice. After that, the teachers could make targeted interventions to improve the effectiveness of teaching and students' learning experience. The purpose of this research is to develop a scale to measure the emotional state of concept mapping and to improve the research on the application of concept maps.

The flexible application of concept maps to solve problems not only requires students to be familiar with this tool but also requires in-depth thinking about the problem. This has a high cognitive load for students, and the increase in cognitive load may lead to emotional collapse of students. Application is also affected by students' emotion inevitably, which indirectly affects their academic performance. To improve students' learning requires teachers to capture students' emotions and to adjust teaching strategies according to students' negative emotional states. However, there is currently a lack of emotional measurement in the process of applying concept maps. The purpose of this research is to develop an emotional scale on the condition of applying concept maps. It aims to lay the foundation for scholars to understand the emotional level of applying concept maps and the influence of emotion on the application of concept maps.

16.3 Method

16.3.1 Research Context

The purpose of this study was to develop the CM-ES and to investigate the emotional experiences of Chinese undergraduate students with concept maps when using them to solve problems.

The study was conducted at a Chinese university that offers a long-standing thinking course, which is a public elective course with students from different majors. In this course, students learned to use many thinking tools, such as Thinking Maps (Hyerle, 2011), Mind Map (Buzan & Buzan, 2010), and Concept Map (Novak, 1984). The course was delivered online by Dingding, which is an online education platform. In that platform, the teachers could arrange online meetings, and students could interact with the instructor at any time during class. In learning concept maps, the teacher first introduced the origin and usage of concept maps. Then the teacher demonstrates how to construct concept maps by CmapTools software.

16.3.2 Participants

A total of 88 college students from a college in China participated in this study. These students came from different majors, including physics, chemistry, history, English, etc. Of all the participants, 86 completed the concept mapping task. Of all participants, 51 filled out the CS-CM questionnaire, and since one of them filled out twice, a total of 52 pieces of data were received. After deleting the duplicate data, the researcher deleted 3 other data that took less than one minute to fill in the questionnaire. Matching the concept mapping task and the CM-ES questionnaire, a total of 48 students completed the whole process, and the data were valid, with a valid response rate of 54.44%. 14.58% ($n = 7$) of them are males and 85.42% ($n = 41$) are females. All students were freshmen.

16.3.3 Data Collection and Analysis

16.3.3.1 Concept Map Task

After students had a basic understanding of concept maps, they were required to create a concept map of the topic of their choice, such as Fig. 16.1. They used the concept map to assist in analyzing the question and responding to the question (Novak & Cañas, 2006). Finally, they were required to submit their concept maps.

16.3.3.2 CM-ES

The purpose of this study was to develop the CM-ES and to investigate the emotional experiences of Chinese undergraduate students with concept maps when using them to solve problems. Immediately after students completed the concept mapping task, they were given a digital version of the CM-ES questionnaire, created by the Questionnaire Star tool, to investigate students' emotions about using the concept maps.

The questionnaire consisted of two parts. The first part of the questionnaire included questions asking for the participants' background information, including major and gender. In the second part of the questionnaire, a total of 26 items adapted from an English emotions scales proposed by Govaerts and Grégoire (2008) were used to examine students' emotion.

The context we conducted the CM-ES is different from the initial scales which is adapted in the situation of a math exam preparation. Hence, all "math" words in the items were replaced with the "concept maps." For example, the item "I am bored studying for the math exam." in the Boredom construct was modified to "I am bored

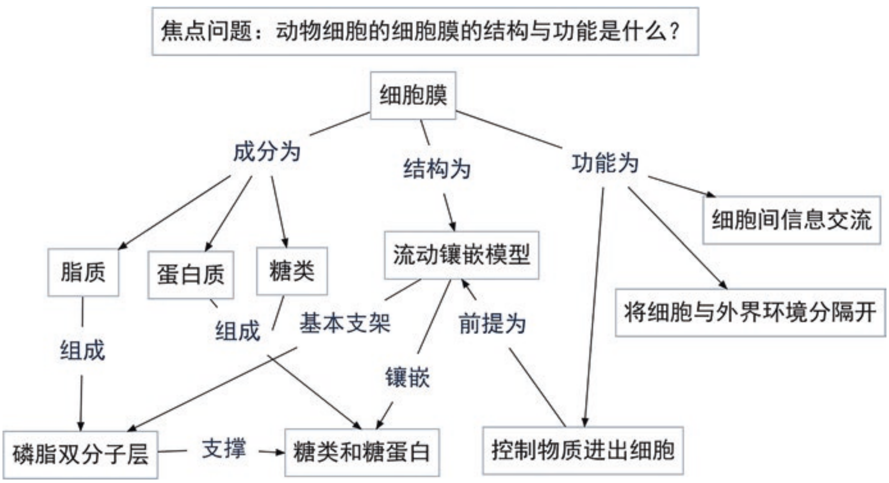


Fig. 16.1 A concept map prepared by a student

studying for the concept maps exam." The item "When I think about the math subjects I have to study, I become anxious" in the Anxiety construct was modified to "When I think about the concept maps, I have to study, I become anxious." To distinguish from the original ones, we add the suffix "CM" to the name of every construct, which means the scales are constructed in the context of drawing concept maps. So, the final names of the six constructs are, respectively, Frustration-CM, Anxiety-CM, Shame-CM, Enjoyment-CM, Hope-CM, and Pride-CM.

The initial CM-ES was designed to measure six components of emotions concerning the construction of concept maps, with six items for Frustration-CM, five items for Anxiety-CM, four items for Shame-CM, four items for Enjoyment-CM, four items for Hope-CM, and three items for Pride-CM. All items were designed on a seven-point Likert scale from 1 "I do not agree at all" to 7 "I totally agree." An educational technology specialist and three postgraduate students majoring in educational technology reviewed the items to improve the readability and content validity of the instrument.

16.3.3.3 Data Analysis

SPSS 25.0 was used to analyze the data. Firstly, exploratory factor analysis (EFA) was conducted with all original items to clarify the structure of the CM-ES, and Cronbach's alphas (α) were calculated to determine the internal consistency of each component of the CM-ES. Secondly, a *t*-test was performed to examine the differences between students of different genders. Lastly, correlation analysis and stepwise regression were implemented to explore the prediction of Frustration-CM, Anxiety-CM, Shame-CM, Enjoyment-CM, Hope-CM, and Pride-CM, respectively, on their CM-ES.

16.4 Findings

16.4.1 Validation of the CM-ES

Exploratory factor analysis with a varimax rotation was used to clarify the structure of the CM-ES survey. Through the first factor analysis (see Table 16.1), the initial items from Anxiety-CM construct, Hope-CM construct, Shame-CM construct, Enjoyment-CM construct, and Frustration-CM construct were all classified into the corresponding components as expected. However, one initial item from the Pride-CM construct was classified into the Hope-CM component, and the other two initial items from the Pride-CM construct were classified into the Enjoyment-CM component. Also, the items Frustration-CM4 and Enjoyment-CM1 were removed as these items did not fit the expected component. After removing these two items, factor analysis was conducted again with these 24 items.

Table 16.1 The first exploratory factor analysis

Item	Factor component				
	1	2	3	4	5
Frustration5	0.760				
Frustration2	0.711				
Frustration1	0.702				
Frustration3	0.699				
Frustration6	0.673				
Frustration4	0.637	0.546			
Anxiety5		0.810			
Anxiety2		0.798			
Anxiety1		0.671			
Anxiety3		0.665			
Anxiety4		0.603			
Hope4			0.867		
Hope2			0.789		
Pride1			0.749		
Hope1			0.747		
Hope3			0.661		
Enjoyment3				0.867	
Enjoyment2				0.789	
Enjoyment4				0.749	
Pride2				0.747	
Pride3				0.661	
Enjoyment1		−0.517		0.523	
Shame1					0.893
Shame2					0.887
Shame3					0.886
Shame4					0.796

The validated CM-ES questionnaire contains 24 items in the 5 new components: Anxiety-CM (five items), Hope-CM (five items), Shame-CM (four items), Enjoyment-CM (five items), and Frustration-CM (five items). Cronbach's α was computed to assess the questionnaire's reliability. Results showed that Cronbach's α of the total scale was 0.765, and Cronbach's α of each dimension (Anxiety-CM, Hope-CM, Shame-CM, Enjoyment-CM, and Frustration-CM) were 0.900, 0.884, 0.908, 0.830, and 0.859, which suggested high reliability.

16.4.2 Students' Scores on the CM-ES

The means and standard deviations of the participants' scores on Anxiety-CM, Hope-CM, Shame-CM, Enjoyment-CM, and Frustration-CM were provided in Table 16.2. It's easy to find that students had positive emotions about using concept

Table 16.2 Descriptive statistics of CM-ES

Item	*N*	Mean	S.D.
Frustration-CM	48	2.52	0.24
Anxiety-CM	48	2.43	0.30
Shame-CM	48	3.16	0.21
Enjoyment-CM	48	3.48	0.19
Hope-CM	48	3.22	0.27

Table 16.3 Gender differences of CM-ES

	Male		Female			
Emotions	Mean	S.D.	Mean	S.D.	*t*	Sig.
Frustration	2.71	0.55	2.48	0.73	0.80	0.43
Anxiety	2.43	0.57	2.43	0.76	−0.02	0.99
Shame	2.79	0.59	3.22	0.97	−1.14	0.26
Enjoyment	3.43	0.68	3.48	0.57	−0.23	0.82
Hope	3.37	0.52	3.19	0.62	0.73	0.47

maps to solve problems. The mean scores for all the dimensions of positive emotions were above 3, the mean scores for the frustration and anxiety dimensions of negative emotions were below 3, and the mean score for the shame dimension was 3.16, which was slightly above 3.

16.4.3 Gender Differences in the CM-ES

As presented in Table 16.3, five emotions showed gender differences. No significant mean differences were detected. However, male students reported more frustration and hope, but less shame than female students.

16.5 Discussion

This study developed and validated a five-dimensional CM-ES to measure students' emotions when using concept maps to analyze problems. The results of the study showed that students had positive emotions on concept mapping, and there were no significant differences between students of different genders on each dimension of the CM-ES.

16.5.1 Instrument Validation of the CM-ES

In this study, the original Pride dimension of the CM-ES was eliminated and ultimately presents five dimensions of the CM-ES. Two of the items in the Pride dimension were classified as enjoyment, and one was classified as hope. This is inconsistent with the findings of Govaerts and Grégoire (2008) for two possible reasons. Firstly, the Pride dimension itself has a smaller number of items of only three compared to the other dimensions. Secondly, Pride, Enjoyment, and Hope are all positive emotions, and different people may interpret and feel emotions differently, and this perception of emotions is influenced by the sample (Terracciano et al., 2003). Thus, it is not strange for emotions containing intrinsic associations to be clustered under the same factor in studies related to academic emotions.

16.5.2 The Emotional State of Freshmen of Concept Mapping

From the results of the study, students had positive emotions about applying concept maps to solve problems. This shows that students do not reject the use of concept maps to solve problems and have a positive attitude toward the results of utilizing concept maps. The frustration and anxiety levels of students using concept maps are relatively low. The reason may be the teacher instructs students to learn enough knowledge and skills to draw concept maps before completing tasks (Brandt et al., 2001). This also indicates that although concept maps can aid cognition, there is still a need to train users in basic knowledge before using them to reduce their negative emotions.

The present study also found no significant emotions difference between male and female students toward the use of concept maps to solve problems. Chiang and Liu (2014) found that in the sample of college students, male students experience stronger emotions in the process of science learning than female students, which is inconsistent with the results of this study. The reason for this phenomenon may be concept maps are learning tools rather than specific learning content. The preference of male and female students for concept maps was not influenced by the learning content, so there was no significant difference in their emotional experience of concept maps. This also facilitates the usage of concept maps by teachers when teaching.

16.5.3 Implications for Practice

The results of this research show that college students' emotions when using concept maps are positive. Therefore, university teachers can use concept maps to teach complex concepts. At the same time, this research suggests that students should first learn about concept maps before utilizing them, which can effectively reduce students' negative emotions toward concept maps. When teachers use concept maps for teaching, they can use the scale developed in this research to measure students' emotions. Then, they could adjust the subsequent teaching strategies for using concept maps according to students' emotions and provide students with necessary scaffolding. After that, the students could make better use of concept maps for learning.

16.5.4 Limitations and Future Work

This study has two limitations as follows. The first limitation of this study is the small sample and the low proportion of male students, so the perception of emotions may be biased by the sample. In future studies, a larger sample size can be used to improve the validity of the questionnaire developed in this study, and further research can be conducted on the effect of emotions on applying concept maps in instructional activities. Also, the emotions measured in this study were static, whereas emotions change dynamically during the students' learning process, and future research could also be conducted with AI technology to dynamically identify the students' emotions when using concept maps to improve further research related to the effect of emotions on the use of concept maps.

16.6 Conclusion

Students have a large number of diverse emotional experiences, both positive and negative, during the learning process (Reinhard Pekrun et al., 2007). These emotional experiences interact with students' motivation and cognition. Therefore, understanding students' emotion when learning is essential (Schutz & Lanehart, 2002). This requires educators to be able to observe students' academic emotions. There is no shortage of scales in psychological research that measure general emotions (Klonsky et al., 2019). However, emotions are contextually influenced, and students' emotional experiences may be completely different in diverse learning contexts (Frijda, 1993). It has been found that general measures of emotion are less predictive of student academic achievement than domain-specific scales (R. Pekrun et al., 2002). This indicates that emotion measures need to be tailored to specific situations to ensure that the results are measured exactly. This study developed and validated the Concept Mapping Emotions Scale, which measures students' emotions when using concept maps to solve problems. With the increasing use of concept maps in K-12 and college education, it is essential to understand students' emotions toward concept maps. When applying concept maps to instruction, teachers can adjust their instructional strategies according to students' emotions and provide necessary scaffolding for students to take greater advantage of the concept maps for learning.

This research focused on the influence of emotions on the use of concept maps and developed and verified a five-construct CM-ES. The five dimensions are confusion, anxiety, shame, enjoyment, and hope, which are used to measure students' emotions of concept mapping. The scale enables researchers to reveal students' emotional experience when using concept maps as a learning tool. Future researchers can conduct researches on how to use concept map to promote deep learning and problem-solving skills of learners. It also provides suggestions for teachers on how to assist students in learning when using concept maps.

The results of the study show that students have an overall positive attitude toward concept mapping. Therefore, it is recommended that teachers use concept maps in the teaching process of complex concepts. However, students need to be trained in concept mapping before applying them, and teachers are supposed to pay attention to students' emotions during concept mapping. Teachers can utilize the scale developed in this research to investigate the emotions of students after concept mapping to provide a basis for the next use of the concept map.

References

Brandt, L., Elen, J., Hellemans, J., Heerman, L., Couwenberg, I., Volckaert, L., & Morisse, H. (2001). The impact of concept-mapping and visualization on the learning of secondary school chemistry students. *International Journal of Science Education, 23*(12), 1303–1313.

Buzan, T., & Buzan, B. (2010). The mind map defined. In J. Harrison (Ed.), *The mind map book: Unlock Your creativity, boost your memory, change your life* (1st ed.). Pearson Education.

Cañas, A. J., Novak, J. D., & Reiska, P. (2015). How good is my concept map? Am I a good Cmapper? *Knowledge Management & E-Learning: An International Journal, 7*(1), 6–19.

Cannon, W. B. (1987). The James-Lange theory of emotions: A critical examination and an alternative theory. By Walter B. Cannon, 1927. *The American Journal of Psychology, 100*(3–4), 567–586.

Chiang, W.-W., & Liu, C.-J. (2014). Scale of academic emotion in science education: Development and validation. *International Journal of Science Education, 36*(6), 908–928. https://doi.org/10.1080/09500693.2013.830233

Chiou, C.-C., Tien, L.-C., & Tang, Y.-C. (2020). Applying structured computer-assisted collaborative concept mapping to flipped classroom for hospitality accounting. *Journal of Hospitality, Leisure, Sport & Tourism Education, 26*, 100243. https://doi.org/10.1016/j.jhlste.2020.100243

Chou, Y.-Y., Wu, P.-F., Huang, C.-Y., Chang, S.-H., Huang, H.-S., Lin, W.-M., & Lin, M.-L. (2021). Effect of digital learning using augmented reality with multidimensional concept map in elementary science course. *The Asia-Pacific Education Researcher.* https://doi.org/10.1007/s40299-021-00580-y

Damasio, A. (2005). *Descartes' error: Emotion, reason, and the human brain (Reprint)*. Penguin Books.

Estrada, C. A., Isen, A. M., & Young, M. J. (1994). Positive affect improves creative problem solving and influences reported source of practice satisfaction in physicians. *Motivation and Emotion, 18*(4), 285–299. https://doi.org/10.1007/BF02856470

Eysenck, M. W., & Calvo, M. G. (1992). Anxiety and performance: The processing efficiency theory. *Cognition and Emotion, 6*(6), 409–434. https://doi.org/10.1080/02699939208409696

Farrokhnia, M., Pijeira-Díaz, H. J., Noroozi, O., & Hatami, J. (2019). Computer-supported collaborative concept mapping: The effects of different instructional designs on conceptual understanding and knowledge co-construction. *Computers & Education, 142*, 103640. https://doi.org/10.1016/j.compedu.2019.103640

Forgas, J. P. (1995). Mood and judgment: The affect infusion model (AIM). *Psychological Bulletin, 117*(1), 39–66. https://doi.org/10.1037/0033-2909.117.1.39

Foster, S., & Lloyd, P. J. (2007). Positive psychology principles applied to consulting psychology at the individual and group level. *Consulting Psychology Journal: Practice and Research, 59*(1), 30–40.

Fredrickson, B. L., & Branigan, C. (2005). Positive emotions broaden the scope of attention and thought-action repertoires. *Cognition and Emotion, 19*(3), 313–332. https://doi.org/10.1080/02699930441000238

Frijda, N. H. (1993). The place of appraisal in emotion. *Cognition and Emotion, 7*(3–4), 357–387. https://doi.org/10.1080/02699939308409193

Furtado, P. G. F., Hirashima, T., & Hayashi, Y. (2019). Reducing cognitive load during closed concept map construction and consequences on reading comprehension and retention. *IEEE Transactions on Learning Technologies, 12*(3), 402–412. https://doi.org/10.1109/TLT.2018.2861744

Gasper, K., & Clore, G. L. (2002). Attending to the big picture: Mood and global versus local processing of visual information. *Psychological Science, 13*(1), 34–40. https://doi.org/10.1111/1467-9280.00406

Götz, T., Zirngibl, A., Pekrun, R., & Hall, N. (2003). *Emotions, learning and achievement from an educational-psychological perspective* (pp. 9–28) https://kops.uni-konstanz.de/handle/123456789/13904

Govaerts, S., & Grégoire, J. (2008). Development and construct validation of an Academic Emotions Scale. *International Journal of Testing, 8*(1), 34–54. https://doi.org/10.1080/15305050701808649

Gracia-Moreno, C., Cerisier, J.-F., Devauchelle, B., Gamboa, F., & Pierrot, L. (2017). Collaborative knowledge building through simultaneous private and public workspaces. In É. Lavoué, H. Drachsler, K. Verbert, J. Broisin, & M. Pérez-Sanagustín (Eds.), *Data driven approaches in digital education* (pp. 553–556). Springer. https://doi.org/10.1007/978-3-319-66610-5_61

Henderson, N., & Milstein, M. M. (1996). *Resiliency in schools: Making it happen for students and educators*. Corwin Press.

Hwang, G.-J., Huang, H., Wang, R.-X., & Zhu, L.-L. (2020). Effects of a concept mapping-based problem-posing approach on students' learning achievements and critical thinking tendency: An application in Classical Chinese learning contexts. *British Journal of Educational Technology*. https://doi.org/10.1111/bjet.13007

Hyerle, D. N. (2011). *Student successes with thinking Maps®: School-based research, results, and models for achievement using visual tools* (2nd ed.). Corwin.

Immelman, S. (2018). *Concept mapping as a technique for scaffolding concept literacy in accounting: The case of an extended programme*. Dissertation, University of Pretoria. https://repository.up.ac.za/handle/2263/68033

Iversen, L., Iversen, S., Dunnett, S., & Bjorklund, A. (2009). *Dopamine handbook* (1st ed.). Oxford University Press.

James, W. (1884). II.—What is an emotion? *Mind, os-IX*(34), 188–205. https://doi.org/10.1093/mind/os-IX.34.188

John, O. P., Robins, R. W., & Pervin, L. A. (Eds.). (2010). *Handbook of personality, third edition: Theory and research* (3rd ed.). The Guilford Press.

Jonassen, D. H. (1992). What are cognitive tools? In P. A. M. Kommers, D. H. Jonassen, J. T. Mayes, & A. Ferreira (Eds.), *Cognitive tools for learning* (pp. 1–6). Springer. https://doi.org/10.1007/978-3-642-77222-1_1

Klonsky, E. D., Victor, S. E., Hibbert, A. S., & Hajcak, G. (2019). The Multidimensional Emotion Questionnaire (MEQ): Rationale and initial psychometric properties. *Journal of Psychopathology and Behavioral Assessment, 41*(3), 409–424. https://doi.org/10.1007/s10862-019-09741-2

Lajoie, S. P., & Derry, S. J. (1993). *Computers as cognitive tools*. Lawrence Erlbaum. (paperback: ISBN-0-8058-1082-X, $32.

Liang, H.-Y., Hsu, T.-Y., & Hwang, G.-J. (2021). Promoting children's inquiry performances in alternate reality games: A mobile concept mapping-based questioning approach. *British Journal of Educational Technology*. https://doi.org/10.1111/bjet.13095

Machado, C. T., & Carvalho, A. A. (2020). Concept mapping: Benefits and challenges in higher education. *The Journal of Continuing Higher Education, 68*(1), 38–53. https://doi.org/10.1080/07377363.2020.1712579

Marzano, R. J., & Kendall, J. S. (2006). *The new taxonomy of educational objectives*. Corwin Press.

Myers, C. D., & Tingley, D. (2017). The Influence of Emotion on Trust. *Political Analysis, 24*(4), 492–500. https://doi.org/10.1093/pan/mpw026

Novak, J. (1984). *Learning how to learn* (1st ed.). Cambridge University Press.

Novak, J. D., & Cañas, A. J. (2006). The theory underlying concept maps and how to construct and use them. *Florida Institute for Human and Machine Cognition, 1*(1), 1–31.

Oaksford, M., Morris, F., Grainger, B., & Williams, J. M. G. (1996). Mood, reasoning, and central executive processes. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*(2), 476–492. https://doi.org/10.1037/0278-7393.22.2.476

Pekrun, R., Goetz, T., Titz, W., & Perry, R. P. (2002). Academic emotions in students' self-regulated learning and achievement: A program of qualitative and quantitative research. *Educational Psychologist, 37*, 91–105.

Pekrun, R., Frenzel, A. C., Goetz, T., & Perry, R. P. (2007). The control-value theory of achievement emotions: An integrative approach to emotions in education. In P. A. Schutz & R. Pekrun (Eds.), *Emotion in education* (pp. 13–36). Academic Press. https://doi.org/10.1016/B978012372545-5/50003-4

Reisenzein, R. (1994). Pleasure-arousal theory and the intensity of emotions. *Journal of Personality and Social Psychology, 67*(3), 525–539. https://doi.org/10.1037/0022-3514.67.3.525

Robinson, M. D., & Clore, G. L. (2002). Episodic and semantic knowledge in emotional self-report: Evidence for two judgment processes. *Journal of Personality and Social Psychology, 83*(1), 198–215. https://doi.org/10.1037/0022-3514.83.1.198

Roessger, K. M., Daley, B. J., & Hafez, D. A. (2018). Effects of teaching concept mapping using practice, feedback, and relational framing. *Learning and Instruction, 54*, 11–21. https://doi.org/10.1016/j.learninstruc.2018.01.011

Roseman, I. J. (1984). Cognitive determinants of emotion: A structural theory. *Review of Personality & Social Psychology, 5*, 11–36.

Rowe, G., Hirsh, J. B., & Anderson, A. K. (2007). Positive affect increases the breadth of attentional selection. *Proceedings of the National Academy of Sciences, 104*(1), 383–388. https://doi.org/10.1073/pnas.0605198104

Scherer, K. R., Schorr, A., & Johnstone, T. (Eds.). (2001). *Appraisal processes in emotion: Theory, methods, research* (1st ed.). Oxford University Press.

Schmitz, T. W., Rosa, E. D., & Anderson, A. K. (2009). Opposing influences of affective state valence on visual cortical encoding. *Journal of Neuroscience, 29*(22), 7199–7207. https://doi.org/10.1523/JNEUROSCI.5387-08.2009

Schutz, P. A., & Lanehart, S. L. (2002). Introduction: Emotions in education. *Educational Psychologist, 37*(2), 67–68. https://doi.org/10.1207/S15326985EP3702_1

Strongman, K. T. (2003). *The psychology of emotion: From everyday life to theory* (5th ed.). Wiley.

Tauber, S. K., Dunlosky, J., Urry, H. L., & Opitz, P. C. (2017). The effects of emotion on younger and older adults' monitoring of learning. *Neuropsychology, Development, and Cognition. Section B, Aging, Neuropsychology and Cognition, 24*(5), 555–574. https://doi.org/10.1080/13825585.2016.1227423

Terracciano, A., McCrae, R. R., Hagemann, D., & Costa, P. T. (2003). Individual difference variables, affective differentiation, and the structures of affect. *Journal of Personality, 71*(5), 669–703.

Tseng, S.-S. (2020). Using concept mapping activities to enhance students' critical thinking skills at a high school in Taiwan. *The Asia-Pacific Education Researcher, 29*(3), 249–256. https://doi.org/10.1007/s40299-019-00474-0

Wang, Y. (2007). On the cognitive processes of human perception with emotions, motivations, and attitudes. *International Journal of Cognitive Informatics and Natural Intelligence (IJCINI), 1*(4), 1–13. https://doi.org/10.4018/jcini.2007100101

Correction to: Examining the Influence of Cognitive Ability on Situating to a Video Game: Expanded Discussion

Michael P. McCreery, Danielle L. Head, Sam A. Leif, Joseph P. Fiorentini, Catherine A. Bacos, Jeffrey R. Laferriere, S. Kathleen Krach, and Le Quanda L. Cole

Correction to:
Chapter 6 in: D. Ifenthaler et al. (eds.), *Orchestration of Learning Environments in the Digital World*, Cognition and Exploratory Learning in the Digital Age,
https://doi.org/10.1007/978-3-030-90944-4_6

The original version of the chapter was inadvertently published with the incorrect spelling of the sixth author's name as Jeffrey R. Lafferiere when it should have been Jeffrey R. Laferriere. This has now been corrected.

The updated version of this chapter can be found at
https://doi.org/10.1007/978-3-030-90944-4_6

D. Ifenthaler et al. (eds.), *Orchestration of Learning Environments in the Digital World*, Cognition and Exploratory Learning in the Digital Age,
https://doi.org/10.1007/978-3-030-90944-4_17

Index

D. Ifenthaler et al. (eds.), *Orchestration of Learning Environments in the Digital World*, Cognition and Exploratory Learning in the Digital Age,
https://doi.org/10.1007/978-3-030-90944-4

S